THE UNIVERSITY *of* EDINBURGH

Development and application of genomic tools to the genetic improvement of Atlantic salmon and Chilean mussels

Carolina Soledad Peñaloza Navarro

A thesis submitted in fulfillment of the requirements for the degree of

Master of Philosophy

The University of Edinburgh

2013

TABLE OF CONTENTS

Declaration		**i**
Acknowledgements		**ii**
Abstract		**iii**
Chapter 1	**General Introduction**	
1.1	**Aquaculture and genetic improvement**	**1**
1.2	**Atlantic salmon and Chilean mussel**	**4**
1.2.1	Salmon	4
1.2.1.1	Salmon genomic resources	6
1.2.2	Bivalves	7
1.2.2.1	Bivalve genomic resources	9
1.2.2.2	The Bivalve Genome	10
1.2.2.3	Genome Inheritance: Heterozygote deficiencies	11
A.	Technical artefacts	13
B.	Population structure	15
C.	Aneuploidy	17
D.	Larval Selection	17
1.3	**DNA Sequencing Technologies**	**20**
1.4	**Thesis Aims**	**24**

Chapter 2 Candidate-gene association study for harvest traits in Atlantic salmon

2.1 Introduction 26

2.2 Material and Methods 28

2.2.1 SNP Identification 28

2.2.2 Association Study 29

2.2.2.1 Animals and Traits 29

2.2.2.2 DNA Extraction and Genotyping 29

2.2.2.3 Statistical Analysis 30

A. Test for Hardy-Weinberg Equilibrium and Linkage Disequilibrium 30

B. Haplotype reconstruction 30

C. Mixed Model association analysis 30

D. Predicted SNP genotype effects 31

2.3 Results 32

2.1.1 SNP Identification and Genotypes 32

2.1.2 Trait Properties 33

2.1.3 Association Study 35

2.4 Discussion 37

Chapter 3 Application of RAD-Seq to discover, genotype and investigate the inheritance of genome-wide SNP markers in the Chilean mussel

3.1 Introduction 41

3.2 Material and Methods 43

3.2.1 Production of F1 Resource families 43

3.2.2	Animals	43
3.2.3	Stage 1: RAD-Sequencing for SNP discovery and characterization	45
3.2.3.1	DNA Extraction	45
3.2.3.2	RAD library preparation and sequencing	46
3.2.3.3	Bioinformatics analysis: RAD-locus assembly and SNP calling	47
3.2.3.3.1	RAD Sequencing terminology	47
3.2.3.3.2	Defining candidate RAD-loci	49
3.2.3.3.3	SNP calling	50
3.2.3.4	Statistical analysis of discovered SNPs	51
3.2.3.4.1	SNP frequency and Genomic distribution	51
3.2.3.4.2	Genome-wide analysis of segregation patterns	52
3.2.4	Stage 2: Investigation of unusual SNP inheritance patterns	52
3.2.4.1	Independent confirmation of RAD-Seq discovered SNPs and haplotypes	52
3.2.4.2	Mixture of mussel samples between families	54
3.2.4.3	Intra-individual tissue variation in genotype	55
3.2.4.4	Extended family testing for Mendelian Inheritance	56
3.3	**Results and Discussion**	**57**
3.3.1	Stage 1: RAD-Sequencing and SNP discovery	57
3.3.1.1	RAD sequencing and representation of the Chilean mussel genome	57
3.3.1.2	SNP discovery in RAD-loci	58
3.3.1.3	SNP frequency in RAD-loci	60
3.3.1.4	SNP distribution: coding vs. noncoding genomic regions	63

3.3.1.5 Genome-wide segregation analysis 63

3.3.2 Stage 2: Investigation of unusual SNP inheritance patterns 67

3.3.2.1 Technical explanations 67

3.3.2.1.1 Independent confirmation of RAD-Seq discovered SNPs and haplotypes 67

3.3.2.1.2 Mixture of mussel samples between families 70

3.3.2.2 Biological explanations 72

3.3.2.2.1 Intra-individual tissue variation in genotype: potential aneuploidy 72

3.3.2.2.2 Extended family testing for Mendelian Inheritance 74

3.4 Conclusions 78

Chapter 4 Heritability of growth-related traits in the Chilean mussel

4.1 Introduction 79

4.2 Material and Methods 81

4.2.1 F1 Resource Families and Phenotypic records 81

4.2.2 Data analysis 82

4.3 Results and Discussion 82

Chapter 5 General Discussions

5.1 Introduction 85

5.1.1 Atlantic Salmon 85

5.1.2 Chilean mussel 87

Bibliography 95

DECLARATION

I declare that this thesis is my own composition and that the research described in it was carried out by me. Specific contributions of others are acknowledged.

Carolina Soledad Peñaloza Navarro

October 2013

ACKNOWLEDGEMENTS

I would like to thank the city of Edinburgh for bringing so many blessings to my life. I would like to thank my supervisors Steve Bishop and Ross Houston for their balanced guidance, which has made me remember how much I like science. I am grateful to my collaborators in Chile, Jorge Toro and Pablo Oyarzun for sending the mussel samples (that was a lot of work). Also I would like to acknowledge Tim Cezard at Genepool who helped with the bioinformatics. Thank you to all the wonderful people in the Institute; especially to Natalie Lowe who taught me everything I know about RAD sequencing and much more. The biggest thank is for Marcelo who shared this path with me, thank you for your infinite patience.

ABSTRACT

Aquaculture is the fastest growing animal food producing sector in the world, providing almost half of the fish consumed worldwide. However, to meet the future large-scale protein demand associated with world population growth, a sustainable increase in production is required. Genetics and genomics techniques have immense potential for enhancing aquaculture production through selective breeding programs, including the incorporation of marker-assisted selection (MAS). These advances are dependent on applying knowledge of the genetic basis of traits of economic importance (*i.e.* their heritability and genetic architecture) and the availability of genomic resources, particularly DNA markers, genome linkage maps and genotyping techniques.

The overall aim of this Master of Philosophy thesis is to investigate the genetic basis of traits of importance to aquaculture, and to develop and characterise genetic markers for potential use in selective breeding. This will be targeted at two aquacultural species of economic importance: the Atlantic salmon (*Salmo salar*) and the Chilean mussel (*Mytilus chilensis*). Since these species are at a very different stage of development of genomic research, with salmon selective breeding and genomics more advanced, species-specific aims were proposed as follows:

1. Atlantic salmon: The objective of this study was to assess the possibility of using gene-specific markers in selective breeding programs by discovering new single-nucleotide polymorphisms (SNPs) in a gene known to regulate growth in mammals and perform a large-scale SNP association study. Novel SNP markers were identified on a gene paralogue (myostatin-1b) that negatively regulates skeletal muscle development and growth. The SNPs were tested for association

with growth and fillet related traits measured in a commercial population of 4,800 Atlantic salmon at harvest.

2. Chilean mussel: The overall aim was to assess the possibility of selective breeding for growth-related traits by assessing their heritability, and by discovering and characterising the inheritance of genetic markers in mussel families. To achieve this, the heritability of mussel weight and shell length at different ages was estimated. In addition, a powerful new method for SNP discovery and genotyping - restriction-site associated DNA (RAD) sequencing - was used to identify markers across the mussel genome, with the specific aims of (i) creating a novel genomic resource, (ii) studying aspects of the genomic architecture, and (iii) examining the inheritance of the markers from parents to offspring, given the unusual inheritance patterns of other marker types described in the bivalve genetics literature.

In the Atlantic salmon population, three novel SNPs were identified on the myostatin-1b gene. One of the SNPs, which was located within the 5' flanking region (g.1086C>T), showed a significant association with several harvest weight traits ($P<0.05$), suggesting an overall effect on fish growth. The SNP acted in an additive matter, with a change from allele C->T associated with an increase in 30 to 50 g in weight depending on the trait.

In the Chilean mussel families, the heritability estimate of body weight was low to moderate (0.11-0.28) and of shell length was not significantly different from zero. These results suggest that selective breeding for body weight is feasible, although environmental factors significantly influence the phenotypic outcome of both the growth-related traits analysed. The analysis of the mussel genome using RAD

sequencing with the SbfI enzyme allowed the discovery of 4,537 putative SNPs. Interestingly, a high SNP frequency was detected in the sequenced mussel genome - one of the highest across metazoans - with an average of 1 SNP per 30 base pairs (bp). In addition, significant distortions from expected Mendelian inheritance ratios were observed in the majority (approximately 70 %) of the discovered SNPs. Finally, frequent presence of non-parental alleles in the mussel offspring was detected. Further experiments were designed to explore potential biological and technical explanations for these phenomena using re-sequencing of selected loci by Sanger technology and including a number of additional mussel families for segregation analysis. The results exclude sequence or genotype error as a technical explanation and validate the high frequency of SNPs, Mendelian distortion and non-parental alleles in multiple families. Therefore, the source of these phenomena is likely to be biological but remains unexplained.

Selective breeding for growth and fillet related traits is routine in salmon production, and the results of this thesis provide an example of a genetic marker in a myostatin paralogue that could advance this selection via the use of MAS. The results presented from the mussel experiment greatly advance genomic resources for this species and provide evidence for heritability of growth traits, thus suggesting that selective breeding is feasible. However, the significant segregation ratio distortions and presence of non-parental alleles in the offspring demonstrate that the inheritance of genetic markers in mussels has yet to be fully understood before MAS can be applied.

AUTHOR CONTRIBUTION

In Chapter 2:

I designed, planned and performed all laboratory and statistical work, apart from the large-scale SNP genotyping (n=4,800), which was contracted to KBiosciences. The output of this Chapter was a paper submitted to BMC Genetics.

In Chapter 3 and 4:

I established collaboration with the Chilean project INNOVA-CORFO "Scientific and Technological development of mussel seeds genetically improved for growth and red tide detoxification rates (Responsible: Marcela Astorga) and imported to the UK tissue from 6 full-sibling mussel families (370 offspring + 8 parents). The Chilean project also provided mussel growth-related phenotypes measured in 22,278 individuals, which I used for estimating heritabilities. I extracted DNA of all the imported mussels and performed the high-throughput RAD sequencing of 96 individuals. SNP discovery and genotyping was performed by Gene Pool (Edinburgh University). I analysed the SNP markers and, because I observed a high number of markers deviating from Mendelian expectations, designed and performed further statistical and laboratory experiments to explore potential explanations.

CHAPTER 1

General Introduction

1.1 Aquaculture and genetic improvement

In the 1960s the "Green Revolution" boosted agricultural developments in a world where the population was growing at near to 2% per annum after decades of improvement in public health. With the promise of feeding 3 billion human beings, science and technology raised world agricultural output by 25 % (Lamb, 2000). With the introduction of new high-yield varieties of rice and wheat, fertilisers, pesticides and agricultural practices, many species and ecosystems were pushed toward the extreme of the curve of production. Despite the hard criticism that has surrounded this phenomenon, it clearly exemplifies the impact - beneficial or not - that technology can make on human nutrition, economic well-being and environmental sustainability. It has been estimated that by 2025 human population will reach 9 billion people. In the face of rapid growing demand for animal protein, and limited land and fresh water resources, aquaculture has emerged as an opportunity to meet future food supply. Furthermore it will also relieve the intense pressure over wild-stock declining due to over-fishing (Brugere and Ridler, 2004). Worldwide aquaculture production has risen from 600,000 tons in 1950 to 52.5 million in 2008 (FAO, 2010). Today science and technology are facing new challenges, and a deep transformation of aquaculture has started in a process known as the "Blue Revolution".

Although the geographical area under aquatic cultivation has expanded substantially over the past 50 years, most of the growth in aquaculture production has been caused

by an increase in productivity driven by new farming practices and technological innovation (Asche and Bjorndal, 2011, Bostock, 2010). It is expected that technological solutions available through advances in engineering and bioscience will continue to improve the efficiency of production. One particular technology that, while proving useful for dramatically improving plant and livestock productivity, has had limited application in aquaculture industry is selective breeding programmes; only 8.2 % of the aquaculture production in 2010 was based on genetically improved stocks (Gjedrem et al., 2012). Selective breeding programs can result in rapid genetic improvement in traits of economic importance, such as farmed animal growth or disease resistance. Prerequisites for effective selective breeding programs include effective tracking and management of families, control of the species' reproduction and phenotypic records for important traits which show additive genetic variation. Additive genetic variation is the component of phenotype (trait) variation that is passed on the next generation, and is quantified by a genetic parameter known as heritability. Heritability (narrow-sense heritability) is the ratio of additive genetic variance (V_A) over the total phenotypic variance (V_P) and is of importance for selective breeding because it is used to evaluate animals and to predict response to selective breeding; a trait with high narrow-sense heritability is expected to respond readily to selection (Falconer and Mackay, 1996). Considering the relatively high heritabilities for economically important traits in fish and shellfish, high phenotypic variability, high fecundities, and short-medium generation intervals (1-4 years), the majority of aquacultured species are good candidates for selective breeding. This is partly reflected by the high genetic gains obtained by aquaculture breeding programs – data on selection response for growth rate in 10 aquatic species average a 14 % increase per generation (Gjedrem and Thodesen, 2005), almost 4 to 5-fold greater than responses obtained for livestock species (Fjalestad et al., 2003). Additionally, as most species used in aquaculture are

not domesticated or are in the early stages of transition towards domestication (*i.e.* adaptation to the culture environment), greater benefits from genetic improvement can potentially be achieved [*e.g.* by efficiently exploiting the high genetic diversity of wild populations; Muir (2005)].

Excluding aquatic plants, 310 species of fish and shellfish were listed by FAO as cultured in 2008 (Bostock et al., 2008). Regarding genetic improvement of fish, selective breeding programs exist for several species, such as Atlantic salmon, Rainbow trout, Nile tilapia and channel catfish. Within the selection goals for these species, long-term and short-term strategies can be identified. Some long-terms strategies utilising family-based breeding programs have achieved 78% increase in weight in Coho salmon after 16 generations (Neely et al., 2008) and a 5-10% increase per generation in gilthead sea bream (Knibb et al., 1997); physiological tolerance to stress in Rainbow trout through the selection for cortisol response traits (Fevolden et al., 2002); resistance to heavy metal (Hg, Cd, and Zn) pollution in Tilapia (Cuvin-Aralar and Aralar, 1995); and an advance in spawning date in Rainbow trout (Siitonen and Gall, 1989). The main short term-strategy exploited for fish is crossbreeding, which in interspecific crosses of Tilapia has translated in higher tolerance to salinity and increase in body weight (Kamal and Mair, 2005). In the case of shellfish genetic improvement, the most advanced development has been achieved for the Pacific oyster *Crassostrea gigas*, for which continuous controlled reproduction in hatcheries has allowed regular large-scale selective breeding programmes for growth and disease resistance in U.S.A, Australia, New Zealand and France.

1.2 Atlantic salmon and Chilean mussel

This thesis aims to develop genomic tools for potential application in the improvement of two species, the Atlantic salmon (*Salmo salar*) and the Chilean mussel (*Mytilus chilensis*). These species are representative of two important aquaculture species groups, the salmonid and the bivalve shellfish. General characteristics of both groups are detailed in the following sections. Further, for bivalves in particular, specific challenges that exist with understanding and interpreting the genome are highlighted and discussed.

1.2.1 Salmon

Salmon are ray-finned fish species that belong to the family Salmonidae. The aquacultured species of salmon occur in two genera, *Salmo* and *Oncorhynchus*. Both genera are endemic to the northern hemisphere, the *Oncorhynchus* to the North Pacific and the genus *Salmo* to the North Atlantic; however they have been introduced to several areas in the world for recreational fishing and aquaculture. The main farmed salmon species are: the Atlantic salmon (*Salmo salar*), with an annual production of 1,236,000 tonnes; the Pacific Coho salmon (*Oncorhynchus kisutch*), with 117,000 tonnes; and the Chinook salmon (*Oncorhynchus tshawytscha*), with a global production of 24,00 tonnes (Gjedrem et al., 2012). The Atlantic salmon is therefore the most important farmed salmonid in the world. Five leading countries account for most Atlantic salmon production: Norway is the dominant producer and exporter (51%), followed by Chile (28%), Scotland (7.4%), Canada (5.7%) and the Faroe Islands (2.7%) (Asche et al., 2013).

Salmon is amongst the most successful aquaculture species when measured by production growth, growing from 299,000 tonnes in 1990 to over 1.9 million tonnes in

2011, at an annual rate exceeding 9.5 per cent (FAO, 2012). This represents an even faster increase than the 7.6 per cent annual growth of total aquaculture production (Gjedrem et al., 2012). There are a number of reasons for this productivity growth, including improved feeds, better production practices at the farms, improved logistics and more efficient supply chains, as well as increased scale of production (Asche and Bjorndal, 2011). Additionally, breeding programmes have been implemented in several countries to improve genetic performance and adaptation to farming environment; to the point at which 97 % of world production is based on improved stock (Gjedrem and Baranski, 2009). A well-characterised example of breeding programmes was the one initiated by the Norwegian government at the beginning of 1970 to improve the Atlantic salmon production. The initial selection for growth performance resulted in offspring of the 5^{th} generation showing a selection response (*i.e.* the change in mean trait value over one generation) of 14 % and a correlated response of 4-5% for feed conversion. Additionally, early sexual maturation – a negative attribute because species divert their energy towards gonad production rather than growth – has reached a selection response of 8% reduction per generation. Partly as a result of selective breeding, production time has been shortened (from 40 months to only 20 months), survival rate has increased and the reliability of production has improved (Ponzoni et al., 2006). In recent years, selective breeding for disease resistance has also received consideration. The moderate to high heritabilities for disease traits obtained by challenging fish populations [*e.g*. h^2=0.43 for bacterial (furunculosis) and h^2=0.32 for viral (infectious salmon anaemia) diseases in the Atlantic salmon; Ødegård et al. (2007)] have motivated the incorporation of disease resistance as breeding objectives.

1.2.1.1 Salmon genomic resources

Considerable basic biological knowledge of salmonids has been developed as a result of their ecological, recreational and commercial value. In addition, since salmonids are undergoing a process of re-diploidisation after a whole genome duplication (WGD) event that occurred 25-100 million years ago (Allendorf and Thorgaard, 1984) they also have been valued as models to study genome rearrangements and speciation processes (Taylor et al., 2001). This general interest has generated research in physiology, nutrition, toxicology, immunology and ecology, among other subject areas (Thorgaard et al., 2002). In this context, a particular emphasis has been given to the development of genetic and genomic resources, particularly for the Atlantic salmon.

Available resources for the Atlantic salmon include cDNA libraries constructed from many tissues and developmental stages and over 500,000 Expressed Sequence Tags (ESTs) (Rise, 2004, Davey et al., 2001, Koop et al., 2008, Andreassen et al., 2009). These ESTs have allowed the development of microarrays for expression analysis and have been used for mining of genetic markers, such as single nucleotide polymorphisms (SNPs) and microsatellites. In addition, sequencing efforts have made publicly available a bacterial artificial chromosome (BAC) library (Thorsen et al., 2005) of a male Atlantic salmon from Norway, which has allowed the construction of a physical map of the salmon genome. As genetic linkage maps are essential tools for marker-assisted selection (MAS), quantitative trait loci (QTL) mapping and functional genomic approaches, several SNP, amplified fragment length polymorphism (AFLP), and microsatellite based linkage maps are available for the Atlantic salmon (Moen et al., 2008, Lien et al., 2011, Gilbey et al., 2004). The need for a large number of genetic markers for genome-wide association studies, fine-mapping and genomic selection, has also promoted the development of SNP-arrays, such as the 6.5 k Illumina iSelect SNP-array developed by the Centre for Integrative Genetics (CIGENE) at Norway (Kent et al.,

2009), and the 130 k Affymetrix Axiom array produced by researchers at Landcatch Natural Selection and the Universities of Edinburgh and Stirling (R. Houston 2013, pers. comm.). Finally, an International Collaboration to Sequence the Atlantic Salmon Genome (ICSASG), representing researchers, funding agencies and industry from Canada, Chile and Norway has been established (Davidson et al., 2010). The first draft assembly (NCBI Assembly GCA_000233375.1) and other salmon genomic resources are publicly available at ASalBase (http://www.asalbase.org).

These genomic resources have been applied to understand the genetic and physiological regulation of traits of importance to the salmon aquaculture industry. While the WGD has resulted in extensive gene duplication, the majority of genes and genetic markers appear to be inherited in a typical disomic fashion. Therefore, techniques and technologies used to elucidate the genomic basis of quantitative traits in other species have been applied to salmonids. For example, markers linked to the QTL affecting resistance to Infectious Pancreatic Necrosis (IPN) have been described, and at present are being used for marker assisted selection (MAS) for disease resistance in Atlantic salmon in Scotland (Houston et al., 2010) and Norway (Moen et al., 2009).

1.2.2 Bivalves

Bivalve shellfish are a class of molluscs comprising more than 30,000 species, including clams, scallops, oyster and mussels. Bivalve species inhabit the temperate and boreal waters of both hemispheres (Gosling, 1992), and can be found in tidal and sub-tidal waters, from rocky coasts to deep-sea hydrothermal vents at >3,000 m depth (Duperron et al., 2006). Most bivalves are filter feeders, extracting suspended organic matter from the water column through active pumping and filtration of water. Filter-

feeding bivalves play an important ecological role in marine and freshwater ecosystems through their influence on benthic–pelagic coupling and nutrient cycling (Newell, 2004). Apart from their ecological role, bivalves are important economic resources for fisheries and aquaculture.

Bivalve shellfish aquaculture represents nearly 21 % (12.9 million tonnes) of world production and 10 % (US$12.8 billion) of world economic value of the aquaculture industry (FAO, 2010). The most important aquacultured species in terms of total bivalve production are clams (36 %) and oysters (35 %), followed by scallops (14.6 %) and mussels (14.4 %). China is the top producer (and also the largest market), responsible for almost 70 % of the world's farmed bivalve shellfish. Indeed, the sharp increase in growth that shellfish industry experienced between 1995 and 2005 – average growth of 5 % per year - was due to China's significant rise in production. Far behind China, the second largest producer of bivalves is Japan, with a production share of 5.8 %. Other important bivalve shellfish producing regions include the United States, South Korea, Thailand, Spain, Chile, New Zealand and Italy (Rees et al., 2010).

Genetic improvements in bivalves have mainly focused on improving growth traits. For oysters the heritability estimates of total weight and shell size have been variable but typically higher than 0.3 (Kong et al., 2013, Li et al., 2011, Wang et al., 2012). Selection based upon these traits can therefore be a successful way of improving growth rate. In the Pacific oyster *Crassostrea gigas* weight yield was improved 9.5 % after one generation of selection (Langdon et al., 2003). In addition, Nell and coll. (1999) selected the progeny of four breeding lines of the Sydney rock oyster *Saccostrea commercialis* for faster growth. The mean weight increased a 4 % and 18 % after one generation and two generations of selection, respectively. Between-generation variation in selection response has also been detected for other oyster species. Newkirk and Haley (1983)

reported a decrease in response to selection for weight in the European flat oyster *Ostrea edulis* in the second generation compared to the first. Similarly, the response for selection on shell width has also been shown to decrease between generations (Wada, 1986). On the other hand, in mussels, heritability estimates for shell length were higher compared to that of oysters, h^2 = 0.8 and 0.5 at 14 and 28 days post-fertilization, respectively (Strömgren and Nielsen, 1989). Shell length selection has been estimated at 24-35% per generation. Likewise, clams exhibit high heritabilities for shell length, ranging from 0.41 at 2 years (Crenshaw Jr et al., 1996) to 0.91 at 9 months of age (Rawson and Hilbish, 1990). In general, improving growth traits in bivalve shellfish through selective breeding appears to be potentially feasible.

1.2.2.1 Bivalve genomic resources

Genomic resources for bivalve species vary according to their relative economic importance. Oyster species have the majority of developed resources, such as BAC libraries (Cunningham et al., 2006), large publicly available EST (Fleury et al., 2009) and cDNA microarray databases (Jenny et al., 2007), linkage maps (Hubert and Hedgecock, 2004, Yu and Guo, 2003, Yaohua et al., 2009), genome-wide (2,782 SNP) markers (Jones et al., 2013), and a whole-genome oyster sequence (Zhang et al., 2012). For mussels, EST databases (Venier et al., 2009) and low-density AFLP linkage maps (Lallias et al., 2007) are the most advanced resources available.

While additive genetic variation clearly exists for traits of economic importance in bivalve shellfish (*e.g.* growth), the development and application of genetic markers in breeding is at a very early stage. Literature on genetic marker inheritance in bivalves frequently suggests atypical patterns of allele frequencies within and across families, which may be reflective of the inheritance of the genome as a whole. An improved

understanding of the inheritance of genetic markers is required before this technology can reliably be applied in breeding for bivalve species. This section of the introduction will review some aspects of bivalves' genomes and also will consider the evidence for unusual genetic marker allele frequencies and segregation patterns, and thus describes some hypotheses relating to the potential causes of these patterns.

1.2.2.2 The Bivalve Genome

Genome size varies greatly across bivalve mollusc. Comparative analysis indicate that C-values [*i.e.* amount in picograms (pg) of DNA contained within a haploid nucleus] ranges between 0.9 pg in the Pacific oyster to 3.4 pg in the horse mussel (Anisimova, 2007), almost 4-fold differences in genome size. Interestingly, apart from this wide interspecific variation, genome-size variation among intraspecific populations of mussels has also been reported, suggesting that a fraction of the bivalve genome is free to vary among individuals of the same species without major biological consequences (Rodriguez-Juiz et al., 1996, Martínez-Lage et al., 1997). Data on standard karyotypes indicate that the majority of mollusc species studied to date are diploid, and that the most frequent chromosome number is 2n = 38 (Thiriot-Quievreux, 2002). Additionally, there is a low correlation between chromosome number and nuclear DNA content. For example, important aquaculture species such as the clam (*Venerupis rhomboides*) and the blue mussel (*Mytilus edulis*) have 38 and 28 diploid chromosomes (Rodriguez-Juiz et al., 1996), respectively, despite showing similar amounts of DNA content. Although Hinegardner (1974) suggested that these quantitative differences between species are due to evolutionary constraints - more generalized molluscs (like mussels) have a higher amount of DNA than specialized species (like oysters) - subsequent evidence

indicates that this relationship is not universal; some cases of evolution through an increase in DNA content have been reported (González-Tizón et al., 2000).

Although bivalve species are under-represented in sequencing projects, the recent *de novo* sequencing and assembly of the Pacific oyster (*Crassostrea gigas*) has provided some early insights into a molluscan genome (Zhang et al., 2012). The genome of the oyster was found to contain abundant repetitive sequences (36 % of the genome) and active transposable elements [discrete pieces of repetitive DNA with the ability to change their position within the genome; Charlesworth and Langley (1989)], confirming similar previous findings in other bivalve species (Saavedra and Bachère, 2006, Kourtidis et al., 2006, Gaffney et al., 2003, Clabby et al., 1996). In addition, the oyster genome revealed a high level of sequence polymorphism rate of 1.3%. Most bivalve species studied to date show remarkable levels of polymorphism. In the Eastern oyster, mining of SNPs from EST databases indicate that of the 6.8 k of sequence analysed, the average density of SNPs was one every 20 bp (Zhang and Guo, 2010). For the Pacific oyster densities of 1 SNP per 60 bp in coding regions and per 40 bp in non-coding regions have been observed (Sauvage et al., 2007). Similarly high polymorphism, about one per 20 bp, were reported in the nematode *Caenorhabditis remanei* (Cutter, 2008) and the sea squirt *Ciona savignyi* (Small et al., 2007). In comparison, the SNP frequency in human genome is about one per 300 bp [Kruglyak and Nickerson (2001) and The International HapMap Consortium (2007)]. Therefore, bivalves are among the species with highest levels of genetic polymorphism.

1.2.2.3 Genome Inheritance: Heterozygote deficiencies

In the 1980s the population genetics of marine bivalves received considerable attention, mainly because the majority of surveyed natural populations exhibited a

lower than expected frequency of heterozygous individuals at multiple marker loci [termed 'heterozygote deficiencies'; reviewed in Zouros and Foltz (1984), Zouros et al. (1988), Gaffney et al. (1990)]. Even though departures from Hardy-Weinberg Equilibrium (HWE) have also been found in fishes, amphibians and reptiles (Waldman and McKinnon, 1993), this high frequency of the observation in marine bivalves sharply contrasted with theoretical expectations. Most marine bivalves display external fertilization followed by a long larval phase, spending weeks to months in the plankton before settling as sessile adults (Thorson, 1950). In addition, species such as oyster, mussels and clams release their gametes in mass-spawning events involving both sexes; Pacific oyster females can produce 23 to 86 million eggs per spawning, while blue mussels may release 7 to 8 million eggs (Davis and Chanley, 1956, Thompson, 1979). These life-history characteristics should result in populations showing (1) HW equilibrium and (2) a high level of genetic homogeneity over wide geographical distances. Despite the frequent observation of genetic homogeneity due to extensive connectivity (mediated by larval transport) in marine organisms (Caley et al., 1996), finding bivalve populations in genetic equilibrium is uncommon. Moreover, heterozygote deficiencies, apart from being extensively observed in bivalve population studies, have also been identified in family-based studies under controlled conditions and where the knowledge of parental genotypes allowed the direct evaluation of expected genotypes in the offspring (Mallet et al., 1985, MacAvoy et al., 2008, Reece et al., 2004), further raising questions about the origin of these deficiencies.

Several hypotheses have been proposed to explain the observed heterozygote deficiencies, ranging from **technical artefacts**, such that heterozygotes are misscored as homozygotes (*e.g.* due to null-alleles); **sampling errors**, due to the incorporation of individuals from differentiated populations or cohorts [*i.e.* spatial or temporal Wahlund effect; Wahlund (1928)]; complex chromosomal rearrangements, which can lead to

aneuploidy; and **selection**, so that heterozygote deficiencies are the by-product of natural selection acting on early larval stages. Each of these theories is discussed briefly below:

A. Technical artefacts

The technical artefact hypothesis postulates that bivalves' deviations from HWE are not a real biological phenomenon, but rather the consequence of particularly high underrepresentation of heterozygotes due to misscoring. The high genetic variability of bivalves is thought to potentially increase misscoring frequency. There are many artefacts in genetic-based analysis that cause misscoring of genotypes, such as allelic-drop out or poor electrophoresis resolution (Miller et al., 2002, Wattier et al., 1998). However, the most common reference in the literature of bivalve genetics is to null alleles. Null alleles are non-amplified (in the case of molecular markers) or non-functional (in the case of biochemical markers) alleles that, when co-segregating with another allele, result in the misscoring of a heterozygous loci as a homozygous one. Null alleles were usually unconfirmed in surveys of wild populations for allozyme variation because information on parental and offspring genotypes are required to demonstrate a null allele. However, although observed in the progeny from pair crosses of pearl oysters, Pacific oysters, blue mussels and dwarf surf clams [see Foltz (1986) and references therein], they were not considered an important source of heterozygote deficiencies mainly because null alleles would be required to be segregating at a high frequency in the populations, implying that bivalves had high mutation rates or high selection coefficients were favouring null heterozygotes, both considered to be unrealistic (Zouros et al., 1980). Following the advent of microsatellite analysis, null alleles resurged as a reasonably strong explanation for heterozygote deficiencies. Since

bivalves are now known to exhibit a high frequency of sequence polymorphisms, primer annealing sites may be affected by point mutations that could prevent primer binding and the subsequent PCR amplification of that allele. Especially critical are point mutations creating primer-template mismatches near the 3' end of the primer sequence, since it has been demonstrated that a single mismatch in this key region can result in complete PCR failure (Kwok et al., 1990).

Null alleles have traditionally been detected in marine bivalve studies. Indeed, the direct assessment of Mendelian ratios in experimental crosses has shown a striking prevalence of null (non-amplifying) alleles. Launey and Hedgecock (2001) evaluated seven Pacific oyster families and found that 15 out of 94 segregation ratios showed genotypes in the progeny inconsistent with parental genotypes; this observation was concluded to be best explained by null alleles. The authors, for example, observed that crosses between a homozygous (AA) and a heterozygous (BC) parent sometimes produced unexpected offspring (BB) or (CC), which led them to suggest that the homozygote parent was actually a heterozygote for a null allele (AØ). The analysis of 79 microsatellite loci using three families derived from crosses among F2 and F3 Pacific oyster hybrids revealed that null-alleles were present at 41 (52%) of the markers (Li et al., 2003). In the same line of evidence, Hedgecock and collaborators (2004) analysed the segregation pattern of 96 microsatellites in three multigenerational families of the Pacific oyster. In total, 49 microsatellites (51%) were reported to have at least one non-amplifying null allele. Notably, despite null alleles being frequently evoked to explain distortions of microsatellite segregation patterns, few studies have directly assessed their presence, for example with primer-redesign or by repeating genotyping assays (Morin et al., 2001, Wasser et al., 2007). As pointed out by Dakin and Avise (2004), when null alleles are suspected in a dataset most researchers simply estimate the frequency of null alleles that would explain the observed deficiency [*e.g*. with the

Brookfield method; (Brookfield, 1996)] and correct the distorted loci; this is also a common practice in population genetics of bivalves. Nevertheless, a study by MacAvoy et al. (2008) provided direct evidence of suspected null alleles by sequencing the region flanking the microsatellite. The authors isolated 49 microsatellites from the endemic New Zealand Greenshell™ mussel to evaluate their potential application in parentage assignment. To confirm that marker polymorphism was due to a change in repeat motif, all microsatellites were sequenced across parental samples. It was concluded that many homozygote genotypes were not identical at the sequence level because of mutation in the region flanking the microsatellite, *i.e.* they were identical by state (allelic size in bp) but not by descent [a.k.a. size homoplasy; Estoup et al. (2002)]. This size homoplasy was extreme for a particular locus (*Pcan1-125*), which showed almost three times as many alleles by sequence as by size. Therefore, individuals appearing as homozygous for alleles at the size level may in fact be heterozygous for alleles at the sequence level.

B. Population structure

In studies of natural populations, sampling errors due to Wahlund effect are also proposed as a potential explanation for heterozygote deficiencies. A Wahlund effect is generated when instead of sampling individuals from a single, randomly fertilized population, a population with an unknown underlying structure is sampled (Wahlund, 1928). Essentially, if two or more subpopulations have independent allele frequencies, the allelic richness parameters will be inflated. Consequently, the occurrence of the "expected" heterozygosity will be restrained, therefore reducing global heterozygosity at marker loci. The most common cause for genetic structuring is geographical barriers to gene flow between the subpopulations, followed by independent genetic drift in each

subpopulation. Although there is generalized notion that marine populations are "opened" due to an apparent lack of physical barriers to gene flow, microgeographical structuring has been detected (Yund and Neil, 2000).

To specifically address the extent to which spatial structuring (microgeographical Wahlund effect) and temporal structuring (temporal Wahlund effect) can account for heterozygote deficiencies, David et al. (1997) sampled 2,855 individuals of the surf clam (*Spisula ovalis*) across three sites (~1km apart). This species displays annual shell lines, allowing circumvention of the difficulty of age estimation. Although significant temporal and microgeographical structuring was detected, it was too weak ($F_{st} \sim 0.001$) to account for the observed heterozygote deficiencies. Johnson and Black (1984) identified a shifting genetic patchiness, which was proposed to be caused by the history of recruitment (temporal Wahlund effect). However, the variation among sites was too small compared to that required to generate observed deficits of heterozygotes. Indeed, despite some observation support the population mixing hypothesis, they are mostly on a macrogeographical scale (Koehn et al., 1976, Johnson and Black, 1984). In reality, the observed heterozygote deficiencies are, in many cases, so large that the required variance in gene frequency among the presumed subpopulations must be larger than the observed in surveys of local populations. Further, heterozygote deficiencies have also been reported in controlled laboratory crosses (Mallet et al., 1985, Reece et al., 2004, MacAvoy et al., 2008) where a Wahlund effect cannot be considered a responsible factor.

C. Aneuploidy

Aneuploidy is a chromosomal abnormality whereby the number of chromosomes in a cell is not an exact multiple of the haploid set (Ricke and Deursen, 2013). Aneuploidy (2n + 1 and 2n - 1) has been reported in ~8% of the embryos of laboratory-spawned mussels, moreover, this frequency rose to 26 % when the mussels were from a hydrocarbon-polluted area (Dixon, 1982). Therefore, chromosomal loss is a plausible biological explanation of the presence of heterozygote deficiency or, conversely, homozygote excess. However, the frequency of aneuploidy is higher in mussel embryos than in adults (Dixon and Flavell, 1986). Since the chromosomal loss exposes potential deleterious mutations previously suppressed by alternative alleles (*i.e.* genetic dominance), aneuploidy is expected to cause high early stage mortalities. At young ages aneuploidy may be contributing to heterozygote deficiencies, however, due to its low frequencies in adult individuals, it is disregarded as a major contributing factor (Corte-Real et al., 1994).

D. Larval Selection

This hypothesis suggests that strong selection against heterozygotes during larval stages followed by a heterozygote advantage after settlement (*i.e.* during their adult sessile phase) could account for deviations from HWE. Selection can be intense in natural populations and cause substantial change over brief time intervals and small geographical scales. Indeed, for bivalves, natural selection is a major driver of genetic change (Hilbish and Koehn, 1985, Pogson, 1991, Sarver et al., 1992); overwhelming the effects of other evolutionary forces such as drift, mutation, and migration. According to the r/K selection theory (Gadgil and Solbrig, 1972), species that evolve under unstable

or unpredictable environments gain little advantage in acquiring adaptive value that permit successful competition with other organisms, and are rather selected for traits that allow to respond opportunistically to favourable conditions. It follows that unpredictable environments will be dominated by species (r-strategists) that have the ability to reproduce rapidly through traits including high fecundity, early maturity onset, small body size, short generation time, and wide offspring dispersion. Bivalves are considered r-strategist as they produce a high number of offspring in every reproductive event. This high fecundity enables higher selection intensity; regardless of intense predation or mortality, at least some of the progeny will survive to reproduce. In bivalves, selection is intense during larval stages (type III survivorship), and mortalities can reach 99 % in various species [in Sprung (1984)]. Larvae suffer mortalities from a variety of sources, for example, from temperature, food quality and predators (Pechenik, 1999). It has also been proposed that larval mortalities are genotype dependant, and that this mortality is sufficient to cause an overall reduction of the number of heterozygotes among populations (Mallet et al., 1985). The hypothesis that selection against heterozygotes during larval stages is responsible for non-Mendelian inheritance of markers is supported by empirical evidence that indicates that the degree of heterozygote deficiency varies with age of sampled bivalves. Tracey and collaborators (1975) examined juvenile and adult populations of the California mussel (*Mytilus californianus*) and found that the magnitude of heterozygote deficiency, despite being observed at both stages, was consistently greater among juveniles than among adults. Likewise, the comparison of different age groups of the eastern oyster *Crassostrea virginica* (from 2-weeks-old to 3-years-old) showed that heterozygote deficiency decreased with age (Zouros et al., 1983). The observation that distortions are larger among younger cohorts is an indication that the deficiency is generated either at fertilization, during the larval stage, or during settlement. Hence, if individuals

were sampled early in their lifetime, before heterozygote loss occurred, minimum deviations from Mendelian segregation (at a family level) or HWE (at a population level) should have been expected. To explore this prediction, Launey and Hedgecock (2001) compared marker segregation ratios in the earliest swimming larval stage (6 hours post fertilization) to those in 2-3 month old juveniles of Pacific oysters. Young larvae showed marker segregation in accordance to Mendelian expectations while juveniles showed distorted segregation ratios. These results suggested that distortions emerged in populations, most likely as a consequence of larval selection. Mallet et al. (1985) provided an additional line of evidence for genotype dependant mortalities by performing an extensive series of pair-matings of the blue mussel *Mytilus edulis*. The comparison of the genotype distribution of progeny at six allozyme markers (age varying from 122 to 337 days) against their parents revealed, apart from the usual multi-locus heterozygote deficiency, that departure from expectations was more similar among related families (sired by the same male) than among unrelated families. Half-sibling families are expected to share more gene variants than unrelated ones, therefore similar deficiencies among related families may indicate shared genotype-dependant selection patterns, further supporting the importance of selection in generating significant deficiencies of heterozygotes.

If selection against heterozygotes in early stages indeed occurs, it would eventually lead to loss of variability, unless the heterozygote state is advantageous in adulthood. Several studies have shown significant correlations between multi-locus heterozygosity and fitness traits in post-larval populations of oyster, mussels, clams and scallops (Garton et al., 1984, Koehn and Gaffney, 1984, Diehl and Koehn, 1985, Pogson and Zouros, 1994); thus, providing evidence that a shift in directional selection during a single-generation could potentially balance losses of variability through increased mortality of heterozygous. Although models of viability selection that may result in

heterozygote deficiency have been proposed (Zouros and Foltz, 1984), the selection against heterozygote larvae still remains counterintuitive.

In summary, to date, deficits of heterozygotes have been observed in a variety of bivalve molluscs, including oysters (Singh, 1982, Singh and Zouros, 1978, Zouros and Foltz, 1983, Saavedra et al., 1993); scallops (Beaumont and Beveridge, 1984); mussels (Koehn and Gaffney, 1984, Diehl and Koehn, 1985, Beaumont, 1991); and several clams, such as Mulinia (Gaffney et al., 1990), Spisula (David et al., 1997), and Ruditapes (Borsa et al., 1991). Although several hypotheses have been proposed to explain this systematic observation, the origin remains unknown. Most likely, there is no single cause to explain deficits in heterozygotes (Gaffney et al., 1990), and each species, locus, and population should be thought of in separate contexts.

Despite the unclear circumstances leading to heterozygote deficiencies, it is essential that the phenomenon is better understood if molecular genetic markers are to be applied in bivalve breeding. It is likely that current high-throughput sequencing technologies can provide a new approach to this longstanding problem.

1.3 DNA Sequencing Technologies

After the foundation of modern genetics was laid by the discovery of the molecular structure of DNA (Watson and Crick, 1953), research into DNA stimulated the development of technologies that allowed the study of biological complexity at a molecular level. As the study of DNA sequence variation is essential for many fields of biological and medical research, sequencing became a backbone technology. Although not the first in the field, the Sanger method (Sanger et al., 1977) is considered the gold standard for sequencing – mainly because it achieves long (>800 bp length), high

quality reads, even for repetitive regions. Sanger sequencing, however, also has several disadvantages. For example, it has a very restricted ability to determine allele frequencies; usually heterozygous SNPs in a PCR product are not represented at a 1:1 ratio, therefore inference of genotypes can become complicated. In addition, Sanger sequencing is cost and labour-intensive, so its use for large-scale sequencing projects (such as whole genome sequencing) is limited. These limitations and the demand for fast, low-cost and accurate genomic information therefore triggered the development of new technologies for sequencing, known collectively as next-generation sequencing (NGS) technologies.

The NGS technologies are different from the Sanger method because they enable millions of sequence reads in one experiment (massive parallel analysis), resulting in a drastically reduced cost and timescale per base of sequence. Three major NGS platforms routinely used are: The 454 Lifesciences (Roche)/GS FLX, which uses pyrosequencing technology; The Illumina/Genome Analyser IIx, which is based on reversible dye-terminators; and The Life Technologies/SOLiD series, which sequences by ligating dye-labelled oligonucleotides. Recently, also single-molecule sequencing has been introduced, namely Single Molecule Real Time (SMRT) and Nanopore Sequencing (Lin et al., 2012). These NGS technologies share aspects of the overall sequencing protocol: they include a template preparation, sequencing and imaging, and data analysis. They however differ in (i) the use of specific protocols for each above-mentioned stage, (ii) the data output, (iii) performance (*e.g.* observed error rate) and (iv) the running costs (Quail et al., 2012). Nevertheless, NGS is an incredibly fast moving field and improvements are constantly increasing performance and decreasing running costs. As an example of the pace of change, the 13-year Human Genome Project completed in 2003 produced the first draft of a whole genome sequence at a cost of

about $2.7 billion. At present, whole genome sequences can be obtained in a few weeks for less than $5,000.

The dramatic shift NGS brought to the world of sequencing was not only epitomized by the ability to produce vast amounts of genetic data cheaply – it meant a democratization of access. For small scale projects on non-model organisms, however, lower sequencing costs still makes collecting whole-genome sequence data for many individuals unaffordable. Moreover, for several studies the purpose of gathering complete genomic sequence data is unnecessary. An alternative approach to use the power of NGS at a smaller, affordable scale is to survey a large set of unlinked genetic markers (rather than the entire genome) spread evenly throughout the genome. Among genetic markers, SNPs represent the most abundant type of variation present in DNA; they are bi-allelic, therefore show simple mutational dynamics; are co-dominantly inherited; and lend themselves to automated allele calling (Landegren et al., 1998, Wang et al., 1998). In virtue of their intrinsic attributes, SNPs have become the markers of choice for variant discovery and genotyping through NGS. To reduce the cost without compromising quality of the SNP calls (which is directly related to sequence depth), several methods have been developed that involve sequencing only a small fraction of the genome for the detection and scoring of markers. One of these methods is Restriction site Associated DNA (RAD) sequencing (from here onwards: RAD-Seq) (Davey et al., 2011, Baird et al., 2008).

RAD-Seq is a powerful method that combines the use of restriction enzymes with NGS (specifically the Illumina platform) for genome-wide marker discovery and genotyping on a population scale. Since it is a NGS-based technique, thousands to hundreds of thousands of genetic markers in tens to hundreds of individuals can be surveyed in a single sequencing step. Briefly, RAD-Seq uses restriction enzymes to produce a reduced

representation (sub-sample) of the genome of multiple individuals. The digested genome of each individual is then ligated to an adapter that has a sequence complementary to the overhang generated by the restriction enzyme (P1 adapter). These adaptors contain a short DNA sequence called a molecular identifier (MID), which allows the pooling of different barcoded DNA samples and therefore tracking of each sample when they are sequenced in the same reaction. Pooled samples ligated to MIDs are randomly sheared to a size suitable for Illumina sequencing, typically 300–700 bp. The pooled samples are then collectively ligated to a P2 adapter. Next, the library (*i.e*. pooled samples) is enriched by PCR to bring the DNA concentrations up to Illumina sequencing requirements; since the P1 and P2 adapters contain the binding sites for the Illumina sequencing primers, only DNA fragments ligated to both adapters will ultimately be amplified by PCR. The enriched library is then sequenced on a high-throughput Illumina sequencer, such as GAIIx or HiSeq 2000 (Baird et al., 2008). Raw RAD Illumina reads are then filtered by applying different quality thresholds or likelihood tests at multiple levels (for example, to raw sequences and SNP calling) (Davey et al., 2013). Finally, filtered sequences are then aligned for genetic variant identification.

RAD-Seq has several applications of particular interest for non-model organism. For example, simulations indicate that it could be feasible to reconstruct phylogenies in younger clades with conserved orthologous restriction sites (Rubin et al., 2012). Also, RAD has been used for fine mapping quantitative trait loci (QTL) (Houston et al., 2012) and identifying loci underlying phenotypic variation or adaptation in natural populations (Hohenlohe et al., 2010). Moreover, for researchers working with organisms lacking a reference genome, the cost-efficient application of RAD-Seq for genome-scale SNPs discovery and genotyping has opened numerous avenues of investigation (Rowe et al., 2011). Until whole-genome (re)sequencing becomes

economically viable for individual laboratories, RAD-Seq technology is expected to continue to be a staple application of NGS to study non-model organisms.

1.4 Thesis Aims

The overall aim of this Master of Philosophy thesis is to investigate the genetic basis of traits of importance to aquaculture, and to develop and characterise genetic markers for potential use in selective breeding. This will be targeted at two aquacultural species of global importance: the Atlantic salmon (*Salmo salar*) and the bivalve mussel (*Mytilus chilensis*). The specific aims are as follows:

Atlantic salmon:

(i) To determine the heritability of growth and fillet-related traits measured at harvest in a large commercial population of Atlantic salmon (*Salmo salar*).

(ii) To investigate the association between polymorphisms within candidate genes for skeletal muscle growth (myostatin-1a and myostatin-1b) and the aforementioned performance traits to evaluate the potential for marker-assisted selection.

Chilean mussel:

(i) To use RAD-Seq to discover SNP markers distributed across the mussel genome and to investigate the inheritance of marker alleles from parents to offspring in full-sibling families.

(ii) To evaluate possible reasons for large-scale SNP allele inheritance distortion [observed in part (i) and in previous literature] exploiting

different sequencing techniques applied to mussels from several additional families.

(iii) To determine the heritability of growth-related traits (weight and shell length) in the Chilean mussel in a large dataset of farmed mussel families.

CHAPTER 2

Candidate-gene association study for harvest traits in Atlantic salmon

---

2.1 Introduction

The identification of genetic markers associated with growth-related traits in Atlantic salmon contributes to the understanding of the physiological regulation of growth and has applications for the genetic improvement of fish stocks. The recent release of genome sequence data for the Atlantic salmon (www.asalbase.org) provides a new opportunity to advance the discovery of sequence variation in candidate genes underlying growth and muscle development traits.

Myostatin (MSTN) is a potent negative regulator of skeletal muscle development and growth in mammals (Lee and McPherron, 1999). Therefore, it has been an attractive candidate gene for the identification of genetic markers for growth and carcass traits in livestock species. Several mutations leading to non-functional MSTN products have been reported to cause the "double-muscling" phenotype characteristic of Piedmontese and Blue Belgian cattle (McPherron and Lee, 1997, Kambadur et al., 1997). Additionally, a SNP in the ovine myostatin gene has been shown to contribute to the muscular hypertrophy of Texel sheep (Clop et al., 2006). The MSTN gene has also been sequenced in lower vertebrates and invertebrates, such as the zhikong scallop, the bighead carp, the Asian sea bass, the shrimp and amphioxus (Hu et al., 2010, Liu et al., 2012, De Santis et al., 2012, De Santis et al., 2011, Xing et al., 2007). However, for the majority of non-mammalian species, the functional role of MSTN is poorly defined.

The phylogenetic analysis of MSTN genes in teleost fish suggested that a whole genome duplication event, which occurred 320-350 million years ago during early fish radiation (Postlethwait et al., 1998, Amores et al., 1998), resulted in two MSTN paralogues being present in modern bony fishes (MSTN-1 and MSTN-2). Additionally, in salmonids, each paralogue was duplicated once again likely due to tetraploidization (25-100 million years ago) (Kerr et al., 2005, Allendorf and Thorgaard, 1984). Thus, a total of four MSTN genes can be found in rainbow trout and Atlantic salmon, namely MSTN-1(-1a and -1b) and MSTN-2(-2a and -2b) paralogues. In contrast to mammals, where the expression of a single MSTN gene is limited primarily to skeletal muscle (McPherron et al., 1997), teleost fish exhibit a broad pattern of differentially expressed MSTN paralogues. For example, in rainbow trout MSTN-1 transcripts have been detected in eye, spleen, testis, brain and muscle, among other tissues (Garikipati et al., 2006). In comparison, MSTN-2 expression pattern has been shown to be more limited and occurs mostly in the brain (Ostbye et al., 2007, Maccatrozzo et al., 2001). While it is unclear exactly what functional role MSTN gene variants play in these tissues, evidence suggest MSTN-1 is involved in muscle development and growth [however, see Gabillard et al. (2013)]. In transgenic zebrafish the suppression of the MSTN-1 gene by RNA interference led to a double-muscle phenotype (Lee et al., 2009), supporting a similar biological function to that previously described in mammals. This functionality is supported by an increase in the number of muscle fibers (hyperplasia) observed in transgenic Medaka that expresses a dominant-negative MSTN (Sawatari et al., 2010).

Since Atlantic salmon MSTN-2b has a frame-shift mutation in exon 1 that leads to a putatively non-functional pseudogene and MSTN-2a is primarily expressed in the brain (Ostbye et al., 2007), then it is reasonable to suggest that MSTN-1b and/or MSTN-1a potentially regulate muscle development in salmon. If true, variation within Atlantic salmon MSTN-1 gene paralogues may be contributing to the phenotypic variation of

economically important harvest traits related to muscle size and growth. Accordingly, relevant polymorphisms could then be applied as tools for marker-assisted selection (MAS) in salmon breeding programmes. The objectives of this Chapter were, first, to investigate the heritability of a set of fillet and growth-related traits measured at harvest; and second, to identify novel SNP markers on the *Salmo salar* (Ss) MSTN-1a and MSTN-1b genes to analyse their association with the aforementioned performance traits to evaluate the potential for MAS.

2.2 Material and Methods

2.2.1 SNP Identification

To discover polymorphism in the Atlantic salmon MSTN-1b and MSTN-1a, specific primer pairs were designed to generate overlapping PCR products from a MSTN-contig that contained the entire gene and ~300bp of flanking sequence at both ends. This contig was built by aligning the complete MSTN-1a gene [Acc. Num. EF392862] and the MSTN-1b gene [Acc. Num. AJ316006.2] with the Atlantic salmon draft genome assembly [Davidson et al. (2010); NCBI assembly GCA_000233375.1]. For SNP discovery and the optimization of PCR amplifications, a panel of ten parental individuals were randomly chosen from randomly-selected families of a commercial population of Atlantic salmon (sourced from Landcatch Natural Selection, Ormsary, UK). By searching for polymorphisms in a subset of the parental samples we may have failed to detect some SNPs in our population, particularly those with a rare allele frequency (*e.g.* < 0.05). However, sampling 10 individuals is adequate to give a good probability (>0.95) of finding SNPs with a minor allele frequency above 0.1. Each PCR product was sequenced using an ABI 3730xl at ARK-Genomics (Roslin, UK) with

forward and reverse primers to check for consistency of sequences. SNPs were identified by visual inspection of both chromatograms with BioEdit (Hall, 1999).

2.2.2 Association Study

2.2.2.1 Animals and Traits

This study was based on 4,800 fish comprising 198 commercial families from Landcatch Natural Selection Ltd. Families were created in 1999 by crossing 136 sires and 198 dams. Pedigree information was available for two previous generations for all individuals with phenotypic records. The phenotypic data was collected at the time of harvesting 3-year old fish and included Harvest Weight (kg), Gutted Weight (kg), Deadheaded Weight (kg), Fillet Weight (kg), Gutted Yield (%) and Fillet Yield (%). In addition, Fat Percentage and Fillet Colour were recorded. Fat Percentage was estimated as the mean of eight readings along the animal's body using the Torry Fatmeter (Distell Ltd). Scores for Fillet Colour ranged from 20-34 units of colour (yellow to red) and derived from the visual contrast of the fillet against the industry standard Roche colour chart [see Powell et al. (2008) for details of trait collection].

2.2.2.2 DNA Extraction and Genotyping

Total genomic DNA was extracted from adipose fin tissue using a Biosprint DNA kit (QUIAGEN, Crawley, UK) following the manufacturer's instruction. The genotyping of the discovered SNPs (see SNP Identification above) was performed on all 4,800 fish by LGC Genomics Ltd (Herts, U.K.) using a KASP assay. The KASP assay is a competitive allele-specific PCR-based genotyping system that allows high levels of assay robustness and accuracy (see technology details at

http://www.kbioscience.co.uk/reagents/KASP.html). Allele-specific primers were designed and utilised by LGC Genomics based on the supplied gene sequences.

2.2.2.3 Statistical Analysis

A. Test for Hardy-Weinberg Equilibrium and Linkage Disequilibrium

Three SNPs segregating in SsMSTN-1b were further tested for departure from Hardy-Weinberg equilibrium (HWE) with a χ^2 goodness of fit test. Linkage Disequilibrium (LD) was estimated for each pair of SNPs, using both the squared correlation, r^2, and the normalised linkage disequilibrium coefficient, D'. Both statistical analyses were performed using Haploview 4.2 (Barrett et al., 2005).

B. Haplotype reconstruction

Haplotypes (*i.e.* the linear combination of adjacent SNP alleles) were inferred for individuals using PHASE 2.1 (Stephens and Donnelly, 2003, Stephens et al., 2001) and used to perform a haplotype-based association analysis.

C. Mixed Model association analysis

Descriptive statistics of harvest traits were performed using SAS software (SAS Inst., Inc., Cary, NC). The heritability values of each phenotype were calculated from the result of the partitioning of variance components (h^2=Va/Vp, where Va and Vp are the additive genetic and the phenotypic variance, respectively) obtained by fitting a single-trait animal model and omitting SNP genotype as a fixed effects in ASReml (Gilmour et al., 2006). To evaluate the relationship between the discovered markers and the

harvest traits, SNP genotypes and haplotypes were included in the model as fixed effects.

The mixed model was as follows:

$Y_{ik} = \mu + G_k + a_i + e_{ik}$

where Y_{ik} is vector of one of the traits on the individual i; μ is the overall mean of the trait; G_k is the fixed effect of the SNP genotype k (3 classes) or the effect of the haplotype (coded as the number of copies given per each haplotype described in the population: 0, 1 or 2 copies); a_i is the additive effect of the i^{th} animal; and e is the residual term. All available pedigree information was included when fitting the model. Statistical significance of the fixed effects was assessed using the Wald F-statistics with denominator degrees of freedom from the fitted model.

D. Predicted SNP genotype effects

For the SNP(s) that showed significant association with a harvest trait, differences between the means of each genotypic class and allelic frequencies were used to estimate additive and dominance effects (Falconer and Mackay, 1996). Standard errors for both effects were calculated from the variance-covariance matrix of the predicted genotype classes and the standard errors of their differences (SED). The percentage of the additive genetic variance (%Va) explained by the SNP was determined using the standard formula: $[2pq\ (a + d\ (q - p))^2]/Va$, where p and q are the major and minor allele frequencies of the SNP, a is the additive effect and d is the calculated dominant effect. The additive genetic variance (Va) was taken from the mixed model without fitting genotype.

2.3 Results

2.1.1 SNP Identification and Genotypes

The Atlantic salmon MSTN-1b and MSTN-1a genes were re-sequenced to screen for polymorphic variation and to assess the association between this variation and harvest traits. Two MSTN-contigs were built and used for primer design. The MSTN-1a contig was built by the alignment of the gene with: Contig_398167, Contig_069377, Contig_141260 and Contig_132711, from the Atlantic salmon draft genome assembly. Whereas the MSTN-1b contig was built by aligning the gene downloaded from NCBI with two contigs of the assembly: Contig_064406 and Contig_398167. After amplification and direct sequencing of PCR products from 10 samples of parental salmon, three SNPs were detected on MSTN-1b and no polymorphisms were detected on MSTN-1a. By re-aligning the PCR products containing the SNPs with SsMSTN-1b, their position in the gene was established: two were located in the upstream putative promoter region of the gene (g.1060C>A and g.1086C>T) and one in the third exon (c.3501C>G). For c.3501C>G, the substitution of alternative alleles was functionally synonymous.

A total of 4,800 animals were genotyped for the three SNPs. Genotypes and frequencies estimated for the three MSTN SNPs are shown in Table 1. The genotype frequencies of g.1060C>A and g.1086C>T were consistent with HWE expectations. However, c.3501C>G was not in HW equilibrium ($P<0.001$), showing a slight deficit of heterozygous animals. The correlations between alleles indicate that g.1060C>A and g.1086C>T are in strong LD (D' 0.99, r^2 0.38). While there is evidence of recombination between g.1086C>T and c.3501C>G (pairwise D' 0.52 and r^2 0.13).

Table 1. Genotype and allele frequencies of SNPs

	g.1060C>A			g.1086C>T			c.3501C>G		
Genotype	**AA**	**AC**	**CC**	**TT**	**TC**	**CC**	**CC**	**CG**	**GG**
Frequency	74	1128	3474	361	1951	2348	3186	1392	102
Minor AF	0.14			0.29			0.17		
Major AF	0.86			0.71			0.83		

2.1.2 Trait Properties

The population of 4,800 commercial Atlantic salmon had been measured for several traits at harvest (approximately 3 years of age). The phenotypic mean, standard deviations and ranges of the traits used in this study are given in Table 3 along with their heritability. All of the weight traits (HWT, GWT, DHWT and FILWT) showed similar and high heritabilities (~0.5), whereas yield traits (GYLD and FILYLD) exhibited low estimates (h^2<0.05). Fat content and flesh colour heritabilities were moderate, viz. 0.17 and 0.29, respectively.

Table 2. Descriptive statistics for harvest weight (HWT), gutted weight (GWT), deadheaded weight (DHWT), fillet weight (FILWT), gutted yield (GYLD), fillet yield (FILYLD), fat content (FAT) and fillet colour (COL) of Atlantic salmon traits. Estimates of heritability (h^2) for each trait and associated standard errors (in parenthesis) are included.

Trait	Mean	SD	Min	Max	h^2
HWT (kg)	2.54	0.63	0.45	7.5	0.52 (0.05)
GWT (kg)	2.32	0.57	0.39	6.8	0.53 (0.05)
DHWT (kg)	2.02	0.51	0.32	6	0.51 (0.05)
FILWT (kg)	1.68	0.42	0.23	5.2	0.53 (0.05)
GYLD (%)	0.92	0.01	0.66	0.99	0.04 (0.01)
FILYLD (%)	0.66	0.04	0.28	1.02	0.05 (0.01)
FAT	12.15	5.64	1.1	41.7	0.16 (0.02)
COL	28.92	0.76	24	33	0.29 (0.16)

2.1.3 Association Study

To assess the association between SNP genotype and harvest traits, a mixed model analysis was performed in 4,759 successfully genotyped individuals. A significant association was observed between g.1086C>T genotypes (P<0.05) and all weight phenotypes (HWT, GWT, DHWT and FILWT) (Table 3), but not with any of the other traits. This significance was maintained when g.1086C>T was fitted either separately or simultaneously in the model with the other SNPs. Genotypes of g.1060C>A and c.3501C>G were not significantly associated with any trait when fitted either individually or simultaneously in the model.

To assess the size of effect associated with the significant MSTN-1b SNP alleles, the predicted trait values for each genotypic class of g.1086C>T were calculated. The thymine allele was associated with an increase in each of the weight traits (P<0.05), with the additive effect of the SNP on these traits ranging from 30 to 50g (Table 3). Whilst this effect was significant, the percentage of the additive genetic variance explained by g.1086C>T in each trait was small and less than 1%. The dominance effect was trivial and non-significant.

Table 3. Genotype means with standard error, additive effect and dominance effects for SNP that showed a significant trait association.

		Genotype Mean ± SE1				
Trait	**SNP**	**TT**	**TC**	**CC**	**a ± SE**	**d ± SE**
HWT	g.1086C>T	2.60± 0.04	2.55± 0.02	2.50± 0.02	0.05±0.017*	-0.0002±0.02
GWT		2.37± 0.03	2.33± 0.02	2.29± 0.02	0.04±0.019*	0.0029±0.02
DHWT		2.07± 0.03	2.04± 0.01	2.00± 0.01	0.03±0.017*	0.0066±0.01
FILWT		1.73± 0.02	1.69± 0.01	1.66± 0.01	0.03±0.014*	-0.0067±0.01

[1]Estimate of the effect is expressed in kg, *P<0.05

Haplotypes were constructed for the three SNPs and the association between haplotype and trait was assessed. In the Atlantic salmon population analysed, 6 haplotypes and 13 diplotypes were identified. Haplotypes with a frequency <0.01 were excluded from further analysis. The most prevalent haplotype, hap5, accounted for 67 % of all haplotypes. The most common diplotype (43 % of samples) was hap5 homozygotes (Table 4). When haplotype combination was fitted in the model, hap4 showed a significant association with the weight traits ($P<0.05$), and the association of hap5 with the same traits was highly significant ($P<0.01$); the exception was fillet weight trait (FWT), for which only hap5 was significant (P-value = 0.031) (Table 5). These two haplotypes differ in a single nucleotide substitution at g.1086C>T, supporting the significant association of this SNP with harvest traits. The occurrence of hap5 at a dosage of two copies in the SsMSTN-1b gene was related with a decrease in 60 to 110g in body weight traits compared to zero copies.

Table 4. Frequencies of common haplotypes and diplotypes.

Haplotype	Frequency	Diplotypes	Frequency
ATG (hap1)	0.02	hap5/hap5	0.43
ATC (hap2)	0.12	hap4/hap5	0.21
CTC (hap4)	0.14	hap2/hap5	0.17
CCG (hap5)	0.67	hap5/hap6	0.07
CCC (hap6)	0.05	hap2/hap4	0.03
		hap4/hap4	0.02
		hap1/hap5	0.02
		hap2/hap6	0.01

Table 5. Haplotype means with standard errors of haplotypes that show significant association with biometrical traits.

	Haplotype 4*± SE[1]			Haplotype 5**± SE[1]		
Trait	**0**	**1**	**2**	**0**	**1**	**2**
HWT	2.50± 0.02	2.56± 0.02	2.48± 0.06	2.60± 0.04	2.53± 0.02	2.49± 0.02
GWT	2.30± 0.02	2.34± 0.02	2.28± 0.05	2.37± 0.03	2.32± 0.02	2.28± 0.02
DHWT	2.01± 0.01	2.04± 0.02	1.98± 0.05	2.08± 0.02	2.02± 0.01	1.99± 0.02
FILWT	1.68± 0.01	1.71± 0.01	1.67± 0.04	1.71± 0.02	1.69± 0.01	1.65± 0.01

[1]Estimate of the effect is expressed in kg, *P<0.05, **P<0.01

2.4 Discussion

In this study, the SsMSTN-1a and SsMSTN-1b genes were re-sequenced for SNP identification. Three novel SNPs were detected on the SsMTSN-1b gene: g.1060C>A and g.1086C>T, both in the 5' flanking region; and c.3501C>G, located in the third exon. The association analysis between the discovered SNPs and eight harvest traits in a commercial population of Atlantic salmon showed that g.1086C>T had a significant association with all weight traits under study (HWT, GWT, DHWT and FILWT). Quantitative trait loci (QTL) associated with flesh colour and growth traits have been described for Atlantic salmon (Baranski et al., 2010, Reid et al., 2005, Boulding et al., 2008, Houston et al., 2009, Gutierrez et al., 2012). However, SsMSTN-1b is linked to markers mapping to chr25 (Ostbye et al., 2007), where only Gutierrez et al. (2012) identified a QTL for body-weight in ~38 month-old fish. In general, QTL scans for growth traits in Atlantic salmon suggest that body-weight traits are highly polygenic. As many locus of small effect are expected to be co-regulating the growth outcome of fish, it is possible that previous QTL mapping studies failed to detect the effect observed in the current study due to a lack of statistical power.

For aquacultural species, a strong association between MSTN polymorphism and production traits has been detected in the bighead carp (*Aristichthys nobilis*), the yellow catfish (*Pelteobagrus fulvidraco*), the spotted halibut (*Verasper variegatus*), the gilthead sea bream (*Sparus aurata*), the common carp (*Cyprinus carpio*) and the Atlantic bay scallop (*Argopecten irradians*) (Liu et al., 2012, Zhu et al., 2012, Li et al., 2012, Sanchez-Ramos et al., 2012, Sun et al., 2012, Guo et al., 2011). Interestingly, for marine species the majority of polymorphisms that have been described to have an effect on weight traits (including the present study) are located on non-coding regions of the MSTN gene. Although this might be a reflection of low levels of coding genetic variants in marine species (De Santis et al., 2012, Liu et al., 2012), it may also be a consequence of the incipient stage of aquacultural research and the shorter history of selective breeding for aquaculture species. For example, it is possible that loss or reduction of function mutations in coding regions of MSTN have not yet been observed and/or selected to appreciable frequencies in commercial aquaculture species. Despite mutations in the coding region of the myostatin gene are known to cause an increase in muscle mass in several mammals, non-coding mutations associated with regulatory pathways may also be underlying phenotypic variation. Moreover, "double-muscling" in Texel sheep is associated with mutations in the 3'UTR of the MSTN gene, which create illegitimate miRNA binding sites and reduces the amount of circulating MSTN protein (Clop et al., 2006). Whether studied marine species encode a more stable MSTN protein and the main effects on growth mediated by this gene are associated with gene regulation rather than gene structure remains to be investigated.

The haplotype-based association analysis was consistent with the significant effect detected for g.1086C>T; the two haplotypes that showed a significant association with weight traits, hap4 and hap5 (P-value<0.05 and P-value<0.01, respectively), differed only by a nucleotide substitution at this locus. In accordance with the predicted trait

values of the genotypes at g.1086C>T, a double dosage of the haplotype carrying the unfavourable SNP, homozygote hap5/hap5, was associated with a decrease in all body weight traits. The difference between the predicted mean of individuals carrying zero or two copies of hap5 varies between 60 g (for FILWT) to 110 g (for HWT).

Myostatin is an important target gene for aquaculture research. Some studies have explored the improvement of growth through the suppression of gene activity by over expressing the MSTN prodomain (Mstnpro). Lee et al. (2010) increased growth rates of rainbow trout by immersing juveniles in bath treatments with flatfish MSTN-1pro expressed in *Escherichia coli*. The improvement of body mass by inhibiting the myostatin gene has also been achieved in African catfish (*Clarias gariepinus*), goldfish (*Carassius auratus*) and tilapia (*Oreochromis aureus*) larvae by an immersion bath treatment with a soluble form of the Active Type IIB receptor (Carpio et al., 2009). However, a study by Kim et al. (2012) showed that the positive growth response achieved by juvenile tilapia under MSTN inhibition by immersion were not sustained until market size; after 45 weeks of exposure to flatfish Mstnpro, no significant weight or length differences between control and immersed tilapia groups was observed.

A promising alternative to short or long-term administration of MSTN inhibiting agents is marker-assisted-selection (MAS) exploiting favourable alleles at naturally segregating polymorphisms in commercial populations of fish. However, in contrast to mammals like the Belgian blue cattle, where a natural deletion of 11 bp on the third exon increased from 20-25 % muscle mass by hyperplasia (Grobet et al., 1997), no genetic variants with considerable impact on weight phenotypes have been identified in fish (Stinckens et al., 2011). Our results show that genetic variation at g.1086C>T has a significant association with growth traits in a commercial population of Atlantic salmon, although the proportion of variance explained by this marker is relatively small

(<1 %). Nonetheless, this SNP should be evaluated further to assess its effect in other populations of salmon and to test any possible functional role on the promoter region of the SsMSTN-1b gene. In other domestic animals, it is likely that polymorphisms with large effects on growth traits are rare due to larger selection pressure moving favourable alleles to fixation quickly (Goddard and Hayes, 2009). However, Atlantic salmon and other aquaculture species are relatively recently domesticated, and polymorphisms of large effect may feasibly still be segregating in commercial populations. Therefore, future studies should aim at evaluating the effect of additional MSTN polymorphisms and other candidate genes for growth in Atlantic salmon. These results would not only be of importance in fish breeding, but also may aid insights into the physiology of muscle growth and development in fish.

CHAPTER 3

Application of RAD-Seq to discover, genotype and investigate the inheritance of genome-wide SNP markers in the Chilean mussel

3.1 Introduction

Molecular markers are the most powerful genomic tools to increase the efficiency and precision of breeding practices for livestock improvement. However, progress in the development of genomic resources for farmed shellfish species, with the notable exception of the Pacific oyster, has been slow. With the advances in NGS, particularly of reduced-representation sequencing methods (such as RAD-Seq), cost-effective, high-throughput SNP discovery is now available for the rapid development of genomic resources in less-studied aquacultured species.

The mussel *Mytilus chilensis* is an important economic resource in Chile. Its aquaculture production has expanded rapidly in the recent years. Since 2005 its production in tonnes has increased ~650 % (from 88,000 to 577,000 tonnes) (SERNAPESCA, 2011). At present, mytiliculture (*i.e.* mussel farming) relies on the yearly collection of wild larval mussels (seeds) during the spawning months. However, larvae show high inter-annual variability in abundance, and this introduces uncertainty in seed supply and damages the industry (Bagnara and Maltrain, 2008). For example, the scarcity of mussel seed in 2012 reduced the production by 30 % compared to the previous year. Moreover, the production in 2013 is expected to be further reduced by at least 40 % because of the same biological factor. In response to this critical problem, hatchery technology is starting to be developed to ensure future seed requirements (Uriarte,

2008). The artificial reproduction and rearing of mussel larvae through hatchery technology would enable a reliable year-round seed supply. Additionally, shellfish hatcheries provide the opportunity to manage breeding, so that commercial populations can be genetically improved for economically important traits such as growth rate, yield, survival and resistance to toxic algal blooms, among others. In this context, because many of the traits are time-consuming or expensive to measure, the development of highly abundant, randomly dispersed SNP markers and their application in mussel breeding has the potential to dramatically reduce the time and cost required for developing genetically improved Chilean mussel.

It is important to remember at this stage (see General Introduction, section 'Genome Inheritance: Heterozygote deficiencies') that mussels - and bivalves in general - present some interesting genetic features that deserve further investigation before the implementation of shellfish breeding programmes or MAS. Primarily, this is related to extensive heterozygote deficiencies [reviewed in Zouros and Foltz (1984); Zouros and collaborators (1988)]. Despite the fact that several non-mutually-exclusive explanations have been proposed to explain this recurrent observation (*e.g*. null-alleles or Wahlund effect), its origin still remains unclear.

The research performed in this chapter occurred in two stages. In stage 1, the aim was to apply RAD-Seq to discover, genotype and characterise the inheritance of genome-wide SNP markers in mussel families in order to create useful genomic resources for potential use in breeding programmes. Stage 1 included investigation of heterozygote deficiencies on a genome-wide scale. In stage 2, the aim was to investigate potential causes of the heterozygote deficiencies with a set of experiments further designed to confirm the validity of the findings of stage 1, and to explore potential sources for these observations at an endogenous, molecular level.

3.2 Material and Methods

3.2.1 Production of F1 Resource families

The mussel families used in this study were provided by the Chilean project INNOVA-CORFO. In 2008 a group of 200 wild mussels (~50 mm shell length) were collected at random from three wild populations from the western coast of South America: Chaihuín (CH) (39°56' S; 73°35' W), Yaldad (YA) (43º 08' S; 73º 44' W) and Punta Arenas (PA) (53°08' S; 70°53' W) (Fig. 1). These individuals were relocated to a hatchery at Puerto Montt (41°30' S; 70°01' W) and allowed to acclimatize for 3 weeks. 150 full-sibling families (~50 families per source location) were created, each by mating two mussels from the same geographical origin. Artificial spawning of the mussels was induced with a thermal cycler procedure; temperature was raised gradually from 15°C to 25°C. The offspring larvae were maintained separately in culture tanks (200 l) and fed once a day with a diet of *Isochrysis galbana* (4 l at an increasing concentration, from 10,000 to 100,000 cel/ml) (Fig. 2). At a juvenile size of ~3 ± 1 mm shell length, F1 progeny were placed in independent, isolated pearl nets (2.5 mm mesh aperture) and transferred to fattening centres located in the Reloncavi estuary. From each family, 1,000 individuals were chosen at random and tagged with unique identifiers.

3.2.2 Animals

Mantle tissue samples (preserved in absolute ethanol) from 6 tagged mussel families were imported to our laboratory for genetic analysis. Each family consisted of the sire, the dam and ~50 offspring aged ~18 months. Two of these families, Family 1 from YA and Family 2 from CH, were used for RAD-Sequencing (Stage 1, see below). The remaining imported mussel families were used for the 'Extended family testing for Mendelian Inheritance' (Stage 2, see below). To develop one of the proposed

experiments for Stage 2, which was designed to explore the potential 'Intra-individual tissue variation in genotype' (see below), 6 individual whole-mussels preserved in absolute ethanol were additionally imported.

All mussel samples used in this study, whether they consisted in mantle tissue or whole-mussels, were imported from the Austral University, Chile, with project-specific import licenses issued by the Scottish Government.

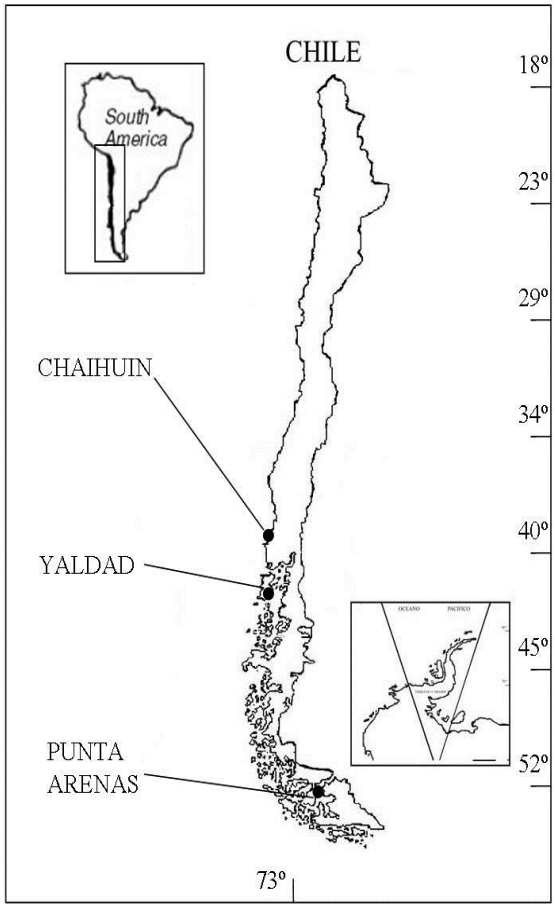


Figure 1. Map of Chile showing locations where the parent stock was sourced: Chaihuín, Yaldad and Punta Arenas.

Figure 2. Larval culture tanks (200 l); each tank contained a single mussel family.

3.2.3 Stage 1: RAD-Sequencing for SNP discovery and characterization

In stage 1, the purpose of the experiments was to discover and characterise SNPs in the mussel genome using RAD-Sequencing.

3.2.3.1 DNA Extraction

Total genomic DNA was extracted from the mantle tissue of the parents and 46 offspring of each of the two families described using a modified CTAB method. Initially, the purity (*e.g.* Absorbance 260/230) and concentration (ng/µl) of the extracted mussel DNA was measured using the Nanodrop™ 800 Spectrophotometer. However, to evaluate the reliability of these DNA measurements, a test RAD library of six mussel individuals was prepared with DNA diluted according to their respective Nanodrop™ measurements. The library was then subjected to PCR amplification with the Illumina sequencing primers to check if the adapters were ligating to the mussel DNA. The PCR failed to yield visible amplified bands on a 1 % agarose gel, meaning that the DNA measurements were significantly overestimated or that the ligation of adapters was

inefficient. To analyse the possibility that an overestimation of DNA was causing non-amplifying libraries, a more sensitive quantification method was evaluated, the PicoGreen™ assay (Invitrogen). This assay exhibits 10,000 more sensitivity than UV absorbance, and is highly selective for double-stranded (ds) DNA in aqueous solution (Singer et al., 1997). A set of mussel DNA samples were measured with both quantification methods (Nanodrop™ and PicoGreen™ assay) and 10-fold differences between methods were detected; Nanodrop™ readings were always higher. Hence – and considering accurate DNA inputs are critical for RAD-Sequencing – the isolated DNA from the two mussel families were ultimately quantified with the PicoGreen™ assay (Invitrogen).

3.2.3.2 RAD library preparation and sequencing

Two full-sibling families - parents and 46 offspring each - were used in the preparation of multiplexed RAD-Seq libraries. 200 ng of dsDNA from each individual was digested with a SbfI high-fidelity restriction enzyme from the New England Biolabs (NEB) for 60 min at 37°C. Samples were heat inactivated for 20 min at 65°C. For ligation of P1 adaptors to the SbfI restriction site, 10nM of SbfI-P1 Adapters containing a 5 bp Molecular Identifier (MID) sequence were added to each sample (see Fig. 3 for adapter details). The ligation reaction was incubated overnight at 16°C. After heat inactivation, 16 samples were pooled and randomly sheared in a Covaris sonicator to an average of 500 bp. The fragmentation conditions used were: duty cycle 10 %, intensity 5, cycle burst 200 and time 105 sec. The sheared, pooled samples were then purified using AMPure beads (Agencourt) at a 1.8:1 ratio of beads to DNA. Fragments ranging from ~250-400 bp were size-selected by agarose gel electrophoresis. A physical gap between libraries (*i.e.* pooled samples) was made in the gel to avoid cross-well contamination. After purification from the gel slice, libraries were end-repaired using

the New England Biolabs (NEB) sample preparation kit (using an incubation time of 30 min). Following end-repair, libraries were purified using a 1.8:1 ratio of AMPure Beads to DNA. A single dA-tail was added to the 3′ end of each blunted, phosphorylated DNA strand using Klenow Fragment DNA Polymerase (3′->5′ exo-) from the NEB sample preparation kit (incubation time of 30 min). At the end of the incubation, the libraries were again purified using a 1.8:1 (Beads/DNA) ratio. Ligation of 10 nM of the P2 adapters was performed at room temperature for 1 hour. All libraries (6 in total) were cleaned with a 0.9:1 (Beads/DNA) ratio and eluted in 50 µL of buffer EB. 5 µL of this product was used in PCR amplification with 50 µL Phusion HF Master Mix (NEB). Phusion PCR settings followed product guidelines for a total of 18 cycles. Libraries were cleaned with 0.9:1 (beads/DNA) ratio, electrophoresed on a 1.1 % gel, and excised between 300–700 bp. Fragments were purified from the gel slice using a MinElute Gel Extraction Kit (Quiagen) and eluted in 20 µL of buffer EB. The six RAD libraries were pooled into a single lane of an Illumina HiSeq 2000 instrument for sequencing at the GenePool Sequencing Facility, University of Edinburgh (http://genepool.bio.ed.ac.uk).

3.2.3.3 Bioinformatics analysis: RAD-locus assembly and SNP calling

3.2.3.3.1 RAD Sequencing terminology

Henceforth, specific terminology for RAD-Sequencing methodology is utilised. Basic terms are defined as follows: a **contig** (from contiguous) represents a consensus region of DNA built by the alignment of a set of overlapping DNA segments. A **read** is an individual raw sequence of a fragment of DNA. Illumina sequencing of short (<1 kb) fragments involves sequencing one or both ends, typically producing reads ranging from 100-150 bp long. If only one end of each ligated DNA fragment is sequenced, specifically the end near the restriction site, this is called the **1st read** or a **single-end read**. On the other hand, if it is chosen to determine the sequence from both extremes

of each DNA fragment, then primary reads (1st reads) and secondary reads (2nd reads) will be obtained, both generating what is called a **paired-end read** (1st read + 2nd read). A **RAD-tag** is a contig (representing either the upstream or downstream DNA region of a SbfI restriction site) built from the multiple alignment of similar primary reads (1st reads). When these RAD-tags are assembled with the consensus sequence of their respective secondary reads (2nd read), a local super-contig or **RAD-locus** is determined. If genetic variability within a RAD-locus is detected, several **RAD-alleles** can be defined, corresponding to individual haplotypes at the locus.

Despite the fact that RAD projects typically utilise only the single-end Illumina sequencing for identification of genetic variants (*e.g.* SNPs or indels), paired-end sequencing of genomic DNA offers several advantages, particularly for studying species with no available reference genome (Davey et al., 2013). For example, 1st and 2nd reads can be assembled into longer contigs (300-600 bp) (Willing et al., 2011) and thus representation of a genome is increased compared to single-end sequencing. Second, paired-end sequencing, in contrast to single-end sequencing, allows the use of information regarding the start and the end of a specific DNA fragment, therefore it can be used to distinguish reads that are derived from independent DNA molecules from those that simply represent PCR duplicates (Baxter et al., 2011). By removing these PCR duplicates not only a data redundancy reduction is achieved, also it mitigates the effect of PCR amplification bias during library construction, which can have a potentially detrimental effect on the quality of SNP calls when coverage is low. Therefore, because of the above mentioned advantages, we used a paired-end RAD sequencing approach to characterise the Chilean mussel samples.

3.2.3.3.2 Defining candidate RAD-loci

SNP calling from the mussel RAD-Seq was performed by the GenePool Bioinformatics Facility, University of Edinburgh (http://genepool.bio.ed.ac.uk). Raw Illumina reads were demultiplexed according to the molecular identifier (MID) assigned to each individual during the RAD library preparation. Each base position in a fluorescence based sequence comes with a quality score (Q) that measures the probability that a base is called incorrectly, the Phred algorithm (Ewing et al., 1998). In this study, reads with a Q <20 (*i.e.* probability of 1 in 100 of an incorrect base call) were removed from the data set by using the process_radtags program of the Stacks package (Catchen et al., 2011).

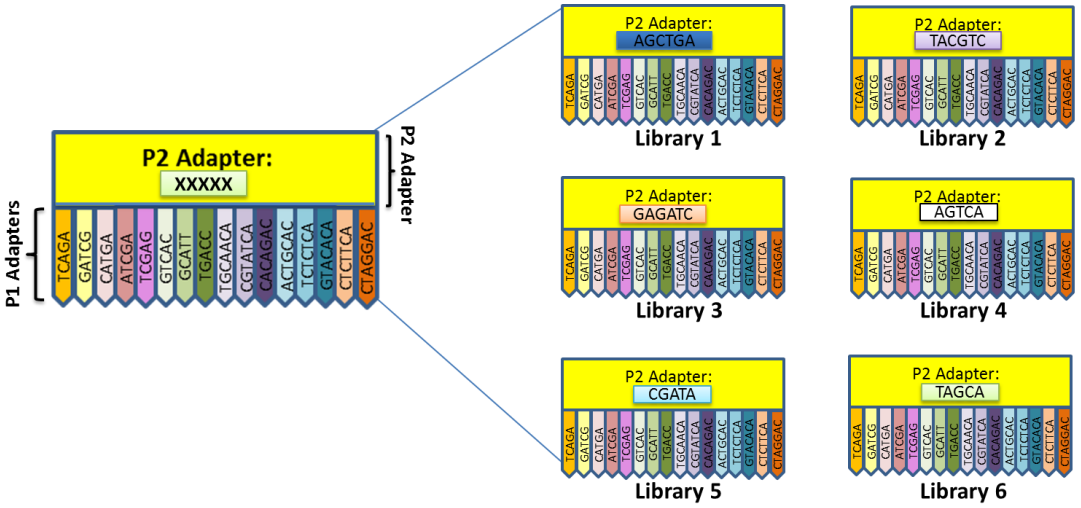


Figure 3. Representation of the multiplexing details of the RAD library preparation. 16 P1-adapters and 6 P2-adapters, each containing a barcode or molecular identifier (MID), are combined in 6 libraries to form 96 samples (*i.e.* individuals) uniquely tagged.

The Chilean mussel lacks a reference genome against which to align sequence reads; therefore a multistep process for *de novo* assembly of RAD-loci was performed. First, the ustack program (Stacks package) (Catchen et al., 2013) was used to identify putative loci within individual mussels. The program formed clusters – also known as "stacks" – of the single-reads allowing a maximum number of 2 mismatches between reads and requiring a minimum depth of 2 reads. A consensus sequence for each stack was determined, defining a RAD-tag. Similarly, the 2nd reads were also clustered and used to build consensus sequences. RAD-tags were then merged across all individuals within a family using cstacks (Stacks package). Afterwards, the RAD-tags and the 2nd reads contigs were assembled into RAD-loci with Merger_EMBOSS (Rice et al., 2000); if both sequences overlapped, a merged consensus sequence was created. Individual genotypes were called by aligning all reads for each individual against the assembled RAD-loci with a minimum genotype quality score of 20.

3.2.3.3.3 SNP calling

Two SNP datasets were derived from the identified mussel RAD-locus, the '*family*' and the '*population*' datasets. The *family-dataset* was created by calling SNPs in the parents and offspring of each family (separately) only if one or both parents were heterozygous for the SNP. Ultimately, only family 1 was used because the RAD sequencing of the dam of family 2 failed. The general characterization of these mussel SNPs markers [*e.g.* segregation (inheritance) analysis] led to the unusual observation that a high number of non-parental alleles were present in the offspring. To further analyse these apparent *de novo* SNPs appearing in the mussel genome, a second dataset was created with the purpose of capturing the majority of genetic variant information contained within each individual mussel. This dataset – the *population-dataset* – was created by calling SNPs

in both families with no imposed constraints on parental genotype; individual mussels were treated as a population rather than consanguineous samples.

3.2.3.4 Statistical analysis of discovered SNPs

3.2.3.4.1 SNP frequency and Genomic distribution

The SNP frequency, defined here as the average distance in bp between SNPs, at each RAD-loci was calculated for both the *family*- and *population*- datasets. The unusually high SNP frequency found in the mussel RAD data (see Results) led to the question of whether there was an association between levels of polymorphism (or SNP frequency) and the genomic location of the consensus RAD-loci sequences, namely if RAD loci belonged to coding or non-coding DNA regions. To examine if SNP frequency differed between putative coding and non-coding sequences, the RAD-loci from the *population-dataset* were used in a BLAST (Altschul et al., 1997) search against the National Centre for Biotechnology Information (NCBI) non-redundant protein database. The rationale was that those RAD-loci that gave significant hits (according to a user-defined significance threshold, the e-value) represented putative coding regions of the mussel DNA, by default the remaining RAD-loci were treated as belonging to putative non-coding DNA. The Genbank non-redundant database is one of the best-annotated sources for comparative *in silico* gene analyses (Benson et al., 2007). However, these records are compromised by a limited number of genes for non-model marine organisms (Cárdenas et al., 2011), therefore limiting the amount of genes that can be queried for significant match. Mussel RAD-loci sequences that showed a significant similarity to annotated proteins using an e-value cut-off of 10-5 were treated as coding sequences, while sequences with no significant similarity to any known protein were considered non-coding sequences. SNP frequencies for the two categories were compared with a t-test.

3.2.3.4.2 Genome-wide analysis of segregation patterns

To analyse the inheritance pattern of SNP markers, RAD data collected on family 1 were used to identify and genotype SNPs from loci that were heterozygous in one or both parents (*family-dataset*); therefore, under standard Mendelian inheritance, ~50 % of the offspring were predicted to be heterozygous and 50 % homozygous at any given locus. Each SNP locus was tested for conformity to expectations of inheritance ratios using a chi-square goodness-of-fit test.

3.2.4 Stage 2: Investigation of unusual SNP inheritance patterns

As outlined in the Results section of this Chapter, the discovered SNPs on the mussel genome showed unusual inheritance patterns when analysed at a family level. Most SNPs markers (72 %) were distorted from the expected segregation ratios and biased towards homozygote genotypes (*i.e*. heterozygote deficient). Additionally, it was common to find genotypes in the progeny (*e.g*. BB) that were not expected from the combination of their progenitor's genotypes (*e.g*. AA x AB). Therefore, a second set of experiments were designed to test various hypotheses that would explain the observed results:

3.2.4.1 Independent confirmation of RAD-Seq discovered SNPs and haplotypes

To validate the accuracy of SNPs and genotypes determined by NGS-based RAD-Seq, a subset of RAD-loci were re-sequenced using Sanger technology and the concordance of results examined. A panel of 34 RAD-loci were used for primer design. Primer pairs were tested in 20 µL PCR reactions containing 200 µM of each forward and reverse primer, 200 µM of dNTPs, 0.04 U of Fast Taq DNA polymerase (Roche), 1 x reaction buffer (2 mM Mg) and 20 ng of mussel genomic DNA. The conditions for amplification of the DNA consisted of an initial denaturation step at 96°C for 1 min and 35 cycles

using the following temperature profile: 96°C for 30 sec, T° annealing for 30 sec, and 72°C for 40 sec. A final extension step at 72°C for 8 min was included. PCR products were electrophoresed on a 2 % agarose gel stained with SYBR Green. PCR products were selected from different test reactions and sequenced to verify reaction specificity. Sequencing was completed by using an ABI 3730xl automated sequencer (Applied Biosystems) at the GenePool Genomics Facility, University of Edinburgh (http://genepool.bio.ed.ac.uk). From the 34 primer pairs tested, *Myt-12396*, *Myt-2828*, and *Myt-9369* were chosen for SNP/haplotype confirmation (Table 1).

The DNA of 5 randomly chosen mussels was amplified with the three selected primer pairs. The purified fragments were subjected to bi-directional Sanger sequencing with the forward and the reverse primer used for the amplification. Visualization of the electropherograms, editing, and alignment of the sequences were performed by using the Lasergene software package (DNASTAR). For SNP validation, specific RAD-loci were displayed on the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013) and compared to the electropherograms obtained for the same genomic region by Sanger sequencing.

Table 1. List of PCR primer pairs used for SNP confirmation

Locus	Sequence 5'-3'	Ta (°C)
Myt-12396	F- AGGAGGCCACCTAGTATCTATCAA	52
	R- ATTGTACCCTTACATCCAATCACA	
Myt-2828	F- CCGAAATTACGTCATTAAAATCAG	52
	R- ATCTCATTGTTCGTACACCAAAAT	
Myt-9369	F- CCGTTGGATCATATAGGGTTGTT	55
	R- GCAAGGGGTAATACCAAAGAATG	

3.2.4.2 Mixture of mussel samples between families

The chance of mixture of mussel samples between families cannot be disregarded as an explanation for (i) extensive heterozygote deficiency of discovered SNP markers, (ii) the non-Mendelian inheritance found in some SNP markers (*i.e*. non-parental alleles in the offspring). Therefore, as a means to examine the validity of the pre-defined family units, the genome-wide relatedness of the 96 individual mussels used for RAD-Seq was examined by calculating a genetic covariance matrix and performing a principal component analysis (PCA) on the SNPs genotypes of the *population-dataset* using the statistical package R (Team, 2011). PCA is a statistical technique than can be used to reduce the dimensions of a multivariate dataset. Specifically, it identifies and quantifies the principal directions in which the data varies. The PCA procedure linearly transforms the original variables into a set of uncorrelated variables, or principal components (PC), which explain essentially the same statistical information (variance) contained in the original data. Each principal component has an associated eigenvalue that measures the respective amount of explained variance. These PCs are ranked according their eigenvalue. As a result, most of the original variability of the dataset is explained by a smaller set of PCs, thus reducing dimensionality. By plotting successive PCs, patterns of genetic structure can be visualized: clusters of individuals can be interpreted as genetic populations, while admixture of two populations results in sets of individuals lying along a line (Lawson et al., 2012). In the case of this study, in the potential absence of major migration between the natural populations from which parental stocks were sourced (CH and YA), the population structure is predicted to reveal two clusters representing each mussel family used for RAD-Seq.

3.2.4.3 Intra-individual tissue variation in genotype

To test the hypothesis that the unusual segregation patterns observed for the SNPs were due to variation in the genotype of the mussels according to which tissue was sampled (*e.g.* due to tissue-specific aneuploidy for example) an intra-individual genetic analysis was performed.

Four whole Chilean mussels were dissected and the gills, mantle, digestive gland, and gonadic tissues (Fig. 6) were extracted. A fraction of the gonadic tissue was used for sexing of the specimens by directly examining the gonad for the presence of sperm or eggs with a stereoscopic microscope. The genomic DNA was isolated using a modified CTAB extraction protocol. DNA quality and concentration of the samples was determined on the NanoDrop™ spectrophotometer.

Primer pairs selected for SNP/haplotype confirmation in mussels (*Myt-12396*, *Myt-2828*, and *Myt-9369*) were used for PCR amplification of specific mussel tissue. Products were direct Sanger sequenced with forward and reverse primers in an ABI 3730xl automated sequencer at GenePool (UK). Electropherograms visualization, sequence editing and within-individual alignments were performed using Lasergene software package (DNASTAR).

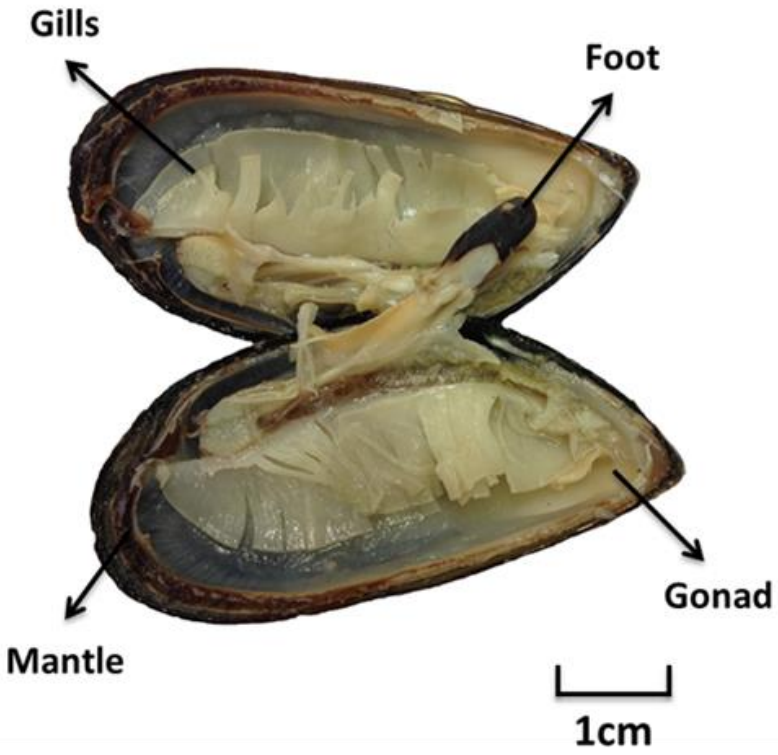


Figure 4. Anatomy of mussel indicating sampled tissues: gills, foot, mantle and gonad.

3.2.4.4 Extended family testing for Mendelian Inheritance

To test that the unusual inheritance pattern was not a feature limited to the two families investigated, the analysis of Mendelian inheritance was extended to four additional mussel families. The genomic DNA of each parent pair and 7 of their respective offspring were PCR amplified with primer pairs *Myt-12396*, *Myt-2828*, and *Myt-9369*. Each PCR product was Sanger sequenced with forward and reverse primers used for the amplification in an ABI 3730xl automated sequencer at GenePool (UK). The generated Sanger reads were aligned within each mussel family using Lasergene software package (DNASTAR) for variant identification.

3.3 Results and Discussion

3.3.1 Stage 1: RAD-Sequencing and SNP discovery

3.3.1.1 RAD sequencing and representation of the Chilean mussel genome

In order to create a novel genomic resource (genome-wide SNP markers), investigate their inheritance patterns, and study aspects of the genomic architecture of mussels, we applied RAD sequencing to discover and genotype SNPs within two full sibling mussel families (each family with two parents and 46 offspring, n = 96). The paired-end sequencing of both families generated 81 million (2 X 100 bp) valid reads, *i.e.* reads with the SbfI restriction site. Since all 96 animals were sequenced within a single Illumina lane, the average number of sequence reads per individual was 0.85 M, although there was a high variability amongst the 96 individuals (σ = 0.45 M).

RAD-Sequencing is a reduced-representation technique where the level of genomic representation (*i.e.* percentage of a total genome sequenced) and depth of sequence coverage per RAD-locus (*i.e*. number of sequenced reads that cover a particular genomic region) is determined by the choice of restriction enzyme, the characteristics of the genome (*e.g*. amount of cytosine and guanine nucleotide content in a genome, a.k.a. GC content) and the number of individual samples multiplexed within a sequencing lane.

In this RAD-project the SbfI restriction enzyme was chosen for the digestion of the mussel genome. The SbfI enzyme has an 8 base, GC-biased recognition site, viz. 5'-CCTGCA^GG-3'. Since the mussel genome has low GC content (~30%) (Tanguy et al., 2008), the expected SbfI restriction site frequency was calculated as follows:

Probability G + C =30% (or 15% each nucleotide = ~1/6.6)

Probability A + T =70% (or 35% each nucleotide = ~1/2.8)

So, the probability of finding a 5'-CCTGCAGG-3' restriction site in the mussel genome is:

$(1/6.6)^6(1/2.8)^2$= 1/648,007 = 1 SbfI restriction site each 648,007 bp

According to the Animal Genome Size Database (http://www.genomesize.com/), mussels have a genome size of 1.56 x 10^9 bp (estimated for *Mytilus edulis* species), therefore ~2,400 potential SbfI cut sites were estimated to be found in the mussel genome. Based on these cut site predictions, ~4,800 RAD-loci (= 2 x n° of genomic SbfI restriction sites) were expected to be obtained after RAD sequencing. In this study, the assembly of RAD-loci in 96 mussel individuals yielded 4,113 RAD-loci. Considering that the average length of these RAD-loci was 300 bp, the RAD sequences covered <1 % of the Chilean mussel genome (genomic size for *Mytilus chilensis* is not available therefore target genome representation was based on estimates of *Mytilus edulis* genomic size). The 4,113 RAD-loci represented an 85 % of the anticipated ~4,800 loci. Observing differences between the numbers of expected to found RAD-loci for a specific genome are common in RAD-Sequencing studies (Hohenlohe et al., 2010, Davey et al., 2013, Mengning Maureen et al., 2013). Some explanations that have been proposed to account for these differences include (i) specific locus effects (*e.g*. PCR bias) that can cause some RAD loci to be sequenced at very low depths, (ii) restriction fragment length bias, (iii) restriction site heterozygosis, (iv) incorrect prediction of enzyme restriction sites, among others (Davey et al., 2013).

3.3.1.2 SNP discovery in RAD-loci

The RAD-Sequence data from the two mussel families were assembled into consensus loci (RAD-loci) and barcoded reads from each individual were aligned to the consensus

for SNP discovery and genotyping. From the 4,113 assembled RAD-loci, 70% (2,903 RAD-loci) contained genetic variation.

There were two separate strategies for SNP discovery and genotyping resulting in two distinct datasets:

In strategy/dataset 1 (defined as the 'family dataset') a set of individual SNP genotypes was generated using a requirement that at least one parent must be heterozygous at a putative SNP position, indicative of SNP segregation within families. Hereafter this is referred to as the 'family dataset'. In family 1, 4,537 SNPs were discovered and genotyped. However, the dam from family 2 had significantly fewer RAD reads than average, causing a very high level of missing genotype data and, consequently, biasing SNP calling within the family. Therefore the application of the SNP calling criteria resulted in the majority of putative SNPs from family 2 being filtered out because of the lack of a reference maternal genotype. Considering the genotypes of the parents of both families were required for a downstream 'Genome-wide segregation analysis' (see below), family 2 was excluded from the family dataset. Visual inspection of the genotype data from this dataset from family 1 revealed that many SNP loci showed non-parental alleles and allele combinations in the offspring.

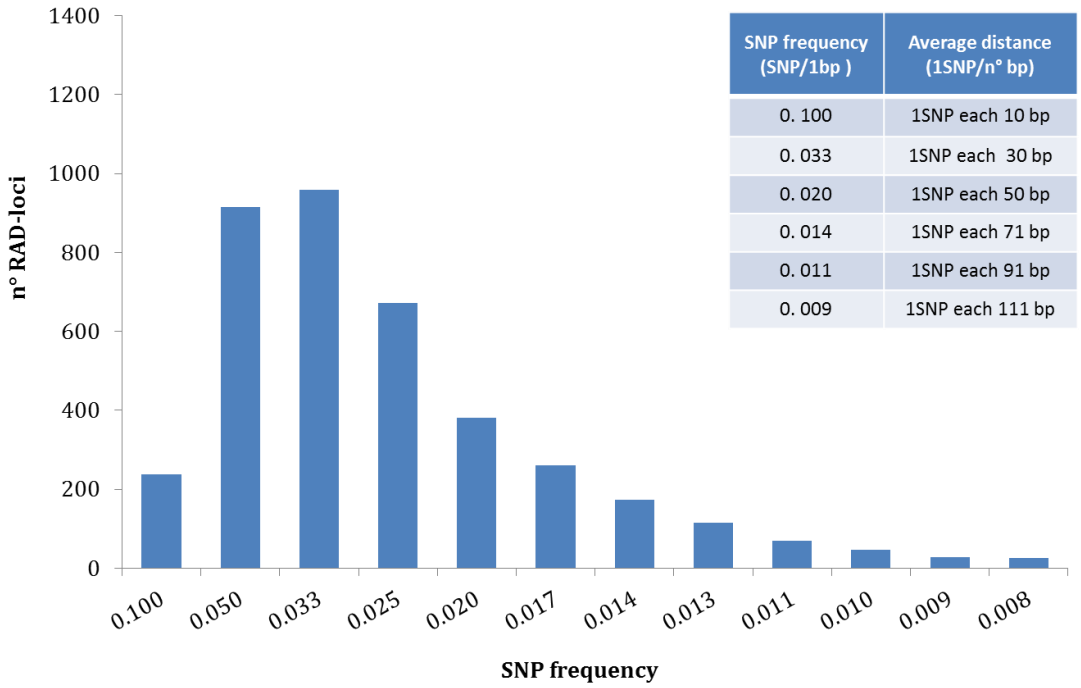
In strategy/dataset 2 (defined as the 'population dataset') SNPs were discovered and called in individuals across both families without the requirement for parental heterozygosity, *i.e.* individual mussels were treated as though they were sampled from a random population rather than consanguineous samples. In the *population-dataset* 65,553 putative SNPs were discovered across both families.

It is noteworthy that not all discovered SNPs in the *family-dataset* and the *population-dataset* were used for further analysis. As explained in the section 'Material and

Methods: Defining candidate RAD-locus', contigs of 1st reads (RAD-tags) and contigs of 2nd reads are assembled into a complete contig defining the flanking region of the SbfI restriction site (RAD-loci). In many instances, when short DNA fragments are abundant, both contigs can overlap and form a 'merged' RAD-locus. In this study, only SNPs discovered from 'merged' RAD-loci were used for analysis because they were assumed to be more reliable. Ultimately, the *family-dataset* comprised 3,490 SNPs distributed throughout 1,471 unique 'merged' RAD-loci; and the *population-dataset* comprised 39,092 SNPs distributed throughout 3,992 unique 'merged' RAD-loci.

3.3.1.3 SNP frequency in RAD-loci

The screening of the RAD-loci revealed the Chilean mussel shows exceptionally high levels of genetic polymorphism. In the *population-dataset*, the average genome-wide SNP frequency observed in the 1,471 RAD-loci was approximately 1 SNP per 29 bp. Whereas in the *family-dataset* the average SNP frequency was 1 SNP per 48 bp. This lower SNP frequency observed in the *family-dataset* is consequence of the more stringent criteria applied for SNP calling, which does not allow capturing the total genetic variation contained within the mussel RAD sequences. A histogram illustrating the distribution of SNP frequencies in RAD-loci (*population-dataset*) reveals a positively skewed distribution, with a mode of 1 SNP per 30 bp (Fig. 4). Similar high SNP frequencies have been observed in other bivalve species. In the Pacific oyster the reported frequency was 1 SNP per 60 bp in coding regions and ~1/40 bp in non-coding regions (Sauvage et al., 2007). Furthermore, an even higher SNP density was estimated by Pante et al. (2012) by transcriptome sequencing of the bivalve mussel *Macoma balthica*, 1 polymorphic site each 19 bp. The high levels of polymorphisms found in the Chilean mussel genome provide evidence supporting bivalves as one of the most polymorphic animal groups studied to date.



SNP frequency (SNP/1bp)	Average distance (1SNP/n° bp)
0. 100	1SNP each 10 bp
0. 033	1SNP each 30 bp
0. 020	1SNP each 50 bp
0. 014	1SNP each 71 bp
0. 011	1SNP each 91 bp
0. 009	1SNP each 111 bp

Figure 4. Distribution of SNP frequencies in the *population-dataset*. Frequency is estimated as the number of SNP per a single bp.

For comparison, SNPs occur in humans every 300-1000 bp along the genome (Sachidanandam et al., 2001). In aquatic animals, average frequency found in the threatened coral *Acropora palmata* was 1 SNP each 257 bp (Polato et al., 2011); in the zebrafish 1 SNP per 219 bp (Woods et al., 2000); and in the California red Abalone, a frequency of 1 SNP each 1000 bp was estimated (Valenzuela-Muñoz et al., 2013). However, similar high SNP frequencies have been observed in some terrestrial plants. For example, in the non-coding genomic region of maize, a frequency of 1 SNP per 31 bp has been estimated (Ching et al., 2002). Also, for a particular species of eucalyptus (*Eucalyptus camaldulensis*), the highest frequency among woody plant species was determined, 1 SNP every 16 bp was (Külheim et al., 2009).

The reasons for the disparity in polymorphism rate between the species is unknown but the interesting shared feature of high SNP frequency found in bivalves and plants may be partly attributed to similarity in their evolutionary biology. Marine bivalves – in common with plants - are highly fecund species; therefore, their germ cells undergo a continuous, large number of cell divisions to produce the amount of gametes required for each reproduction event (typically 10^6-10^8 eggs per female per season in the Pacific oyster). It has been observed that mutations result from errors in DNA replication [reviewed in Campbell and Eichler (2013)]. As a consequence, the more germ line divisions (DNA replication), the higher the expected mutation rates. This suggests that the high SNP frequency found in the mussel genome, and other bivalve species as well, might be caused by elevated mutation rates due to high fecundity. These highly fecund species create numerous genetically diverse offspring each reproduction event. In the context of natural selection, the strategy of generating a high numbers of genetically polymorphic offspring may be explained because bivalves inhabit (and evolve) in extremely variable environments. Therefore, the strategy of survival is placed not on parents transmitting well-adapted genotypes to the offspring, but on parents producing many offspring, which in addition to being genetically variable because of sexual reproduction (and underlying genetic recombination) increase their genetic variability through germ line mutations. These abundant, highly polymorphic offspring may permit a large amount of natural selection to take place in a single generation; mortalities over the lifetime of a cohort are progressive and can reach > 90 %. However, because of bivalves' high fecundities the survival of a few, better-fit individuals are apparently sufficient to ensure the maintenance of this group of species. Indeed, if a local habitat is likely to change in time fitness will have virtually zero heritability, therefore stocking the environment with a high diversity of individuals is likely to provide better chances of species survival.

3.3.1.4 SNP distribution: coding vs. noncoding genomic regions

Many mutations are generated essentially randomly in the DNA of individuals. As a result, coding and noncoding DNA are almost equally susceptible to mutation. However, mutations in coding regions are likely to have a deleterious effect on gene expression and so can result in disease or lethality. Thus selection pressures reduce the overall frequency of surviving mutations in coding DNA. This leads to coding regions showing lower SNP frequencies than noncoding genomic regions (Strachan and Read, 1999).

Since RAD-Seq methodology is based on the non-specific digestion of genomic DNA with a restriction enzyme. RAD-loci, and therefore derived RAD markers (SNPs) are anonymous (*i.e.* their location within the genome is unknown). In this study, the analysis of the RAD-loci revealed that the mussel genome shows high levels of SNP polymorphisms. To further characterize this interesting high SNP frequency, we sought to determine if there was an underlying association between the nucleotide diversity of RAD-loci and their correspondence with functional categories of DNA, namely coding and non-coding regions. To achieve this, RAD-loci from the *population-dataset* were aligned using BLASTX against non-redundant protein databases to distinguish putative coding from noncoding sequences. Of the 3,992 RAD-loci sequences identified in the *population-dataset*, 480 retrieved significant hits. The majority of the hits (~52 %) correspond to Pacific oyster proteins. The results from the t-test indicate that there is no significant difference between the SNP frequency found in putative coding and non-coding RAD-loci sequences (P-value = 0.12).

3.3.1.5 Genome-wide segregation analysis

In the family dataset, we evaluated the offspring segregation ratios for all the discovered RAD SNPs to test for deviations from the expected ratios according to

Mendel's laws. A widespread observation in the literature of bivalve genetics is molecular markers showing deficiency of heterozygous genotypes relative to (i) Hardy-Weinberg expectations (in population studies) or (ii) Mendelian segregation expectations (in family studies). Another interesting observation is that Mendelian segregation analysis within bivalve families revealed that while some markers are shown to segregate from parents to offspring following Mendel's law (Reece et al., 2004), at other loci alleles are apparently not being inherited in accordance to Mendelian expectations (MacAvoy et al., 2008). This unusual inheritance of alleles comes in two forms: the offspring show non-parental alleles (*e.g*. AA x AB = AC); or the offspring effectively show the parental alleles, but however display combinations (genotypes) that disagree with predictions from parental genotypes (*e.g.* AA x AB = BB), suggesting that some genetic information in bivalves is non-Mendelian. However, both general observations - heterozygote deficiencies and non-Mendelian inheritance of genetic markers – are mostly based on microsatellite analysis, which are prone to allele mistyping and therefore genotype misscoring. Furthermore, because of the high genetic polymorphism of bivalves, the prevalence of typing errors derived, for example, from null (non-amplifying) microsatellite alleles or microsatellite size homoplasy [where different alleles are identical in size but not identical by descent Estoup et al. (2002)] are expected to be higher, thereby potentially undermining conclusions derived from genetic studies on bivalve that used this marker type. In addition, the majority of these studies used a limited number of microsatellite markers [typically < 15; *e.g.* Reece et al. (2004) and MacAvoy et al. (2008)], thus sampling a small genomic region. To gain a better understanding of the (apparent) unusual genetic inheritance of bivalves along with their heterozygote deficiencies, genome-wide surveys of genetic variation and their transmission in addition to accurate genotyping methods are required. Here, RAD-Seq not only was applied to provide a novel genomic resource (*i.e.* genome-wide

SNP markers), but also was used to improve the framework in which to study inheritance patterns and heterozygote deficiencies. By analysing pedigreed mussel families and particularly by directly identifying their genetic variability at a sequence level – thus at least partly removing technical artefacts such as null or homoplastic alleles – bivalve's unusual genetic features were studied, for the first time, with high marker reliability at an unprecedented genome-wide scale.

The discovery of 4,537 SNPs on the *family-dataset* provided the opportunity to examine the genome-wide inheritance pattern of SNP markers as a means to assess the extent to which heterozygote deficiencies were common features of the mussel genome. According to the criteria for SNP calling utilized to create the *family-dataset* (see Material and Methods), if one or both parents are heterozygous for a putative SNP, the expectation under normal Mendelian inheritance is for 50 % of the offspring to be heterozygous at the locus and 50 % to be homozygous. Of the total 4,537 SNPs in family 1, 3,280 (72 %) markers showed a deviation from Mendelian expectations. The majority of these distortions indicate a deficiency of heterozygotes in the progeny (Fig. 5). Additionally, the *family-dataset* was subjected to a 'Mendelian error' evaluation, where each mussel trio (comprising the sire, the dam and a particular offspring of the family) was evaluated for conformity to Mendelian inheritance for each discovered SNP locus. The results indicate that in family 1 1,909 Mendelian errors (*i.e*. distortions in inheritance pattern) were found.

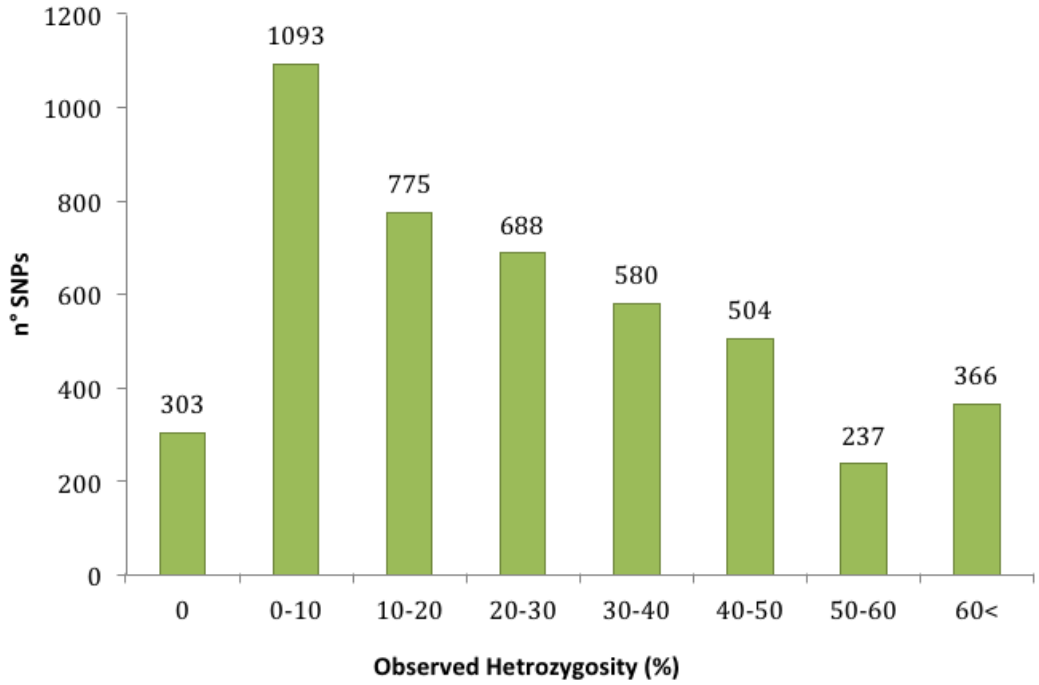


Figure 5. Histogram indicating SNP heterozygosity observed in the *family-dataset*.

The overall observed deficiency in heterozygote genotypes was surprisingly high, indicating that more than 70% of the discovered SNPs deviated from Mendelian segregation expectations. These results are in agreement with studies that detected heterozygote deficiencies by using allozyme and microsatellites as genetic markers, however, our results additionally reveal that deficiencies are extensively dispersed throughout the mussel genome. Moreover, the importance of these results lies in me then being able to address one of the most frequent hypotheses proposed to explain recurrent heterozygote deficiencies, viz. that these are a of technical nature, specifically caused by null-alleles (Panova et al., 2008). However, by using RAD-Seq, markers from the mussel genome were determined directly from the sequence of the RAD-locus, thus removing null-alleles as a significant source for deficiencies. The genome-wide

segregation analysis suggests that, despite removing potentially significant technical sources of heterozygote deficiencies, the magnitude of the phenomena still remains, hence indicating that heterozygote deficiencies in bivalves might be of true biological origin.

3.3.2 Stage 2: Investigation of unusual SNP inheritance patterns

Since a very high SNP frequency, significant genome-wide distortions from Mendelian inheritance, and frequent instances of non-parental alleles in the offspring were detected, further experiments were designed to explore potential explanations for these phenomena. Both technical and biological possibilities were investigated.

3.3.2.1 Technical explanations:

3.3.2.1.1 Independent confirmation of RAD-Seq discovered SNPs and haplotypes

In this study SNP discovery and genotyping of mussel families was achieved using NGS-based RAD-Sequencing technology. NGS generates hundreds of millions of short (30–100 bp) reads, which, in the absence of a reference genome, have to be assembled to reconstruct the original genomic sequence of the sample(s) (*i.e*. *de novo* assembly). The *de novo* assembly of highly polymorphic genomes with abundant number of repeats can be challenging (Kumar et al., 2012).

In order to examine if the assembled RAD-loci reflected the true genetic variability of mussels, a subset of the discovered SNPs were validated by Sanger sequencing. Three primer pairs spanning 3 different RAD-loci were chosen for SNP confirmation: *Myt-12396*, *Myt-2828*, and *Myt-9369*. Since the loci amplified by these primer pairs represent genomic regions with a high number of single-nucleotide variants (SNPs),

over 40 genotype calls were tested. The correspondence between the Sanger sequences and the RAD-loci sequences were high, which indicates that the results obtained using the Illumina HiSeq 2000 platform and the RAD SNP calling protocol are reliable (see Fig. 6). However, the analysis of the genomic region sampled by the primer pair *Myt-9369* revealed discrepancies between both sequencing platforms. A region supposed to span eight heterozygote SNPs in two individual mussels – data from the RAD-Seq reads – is seen fully homozygous in the Sanger electropherograms. Examples of SNP validations are presented in Fig.6.

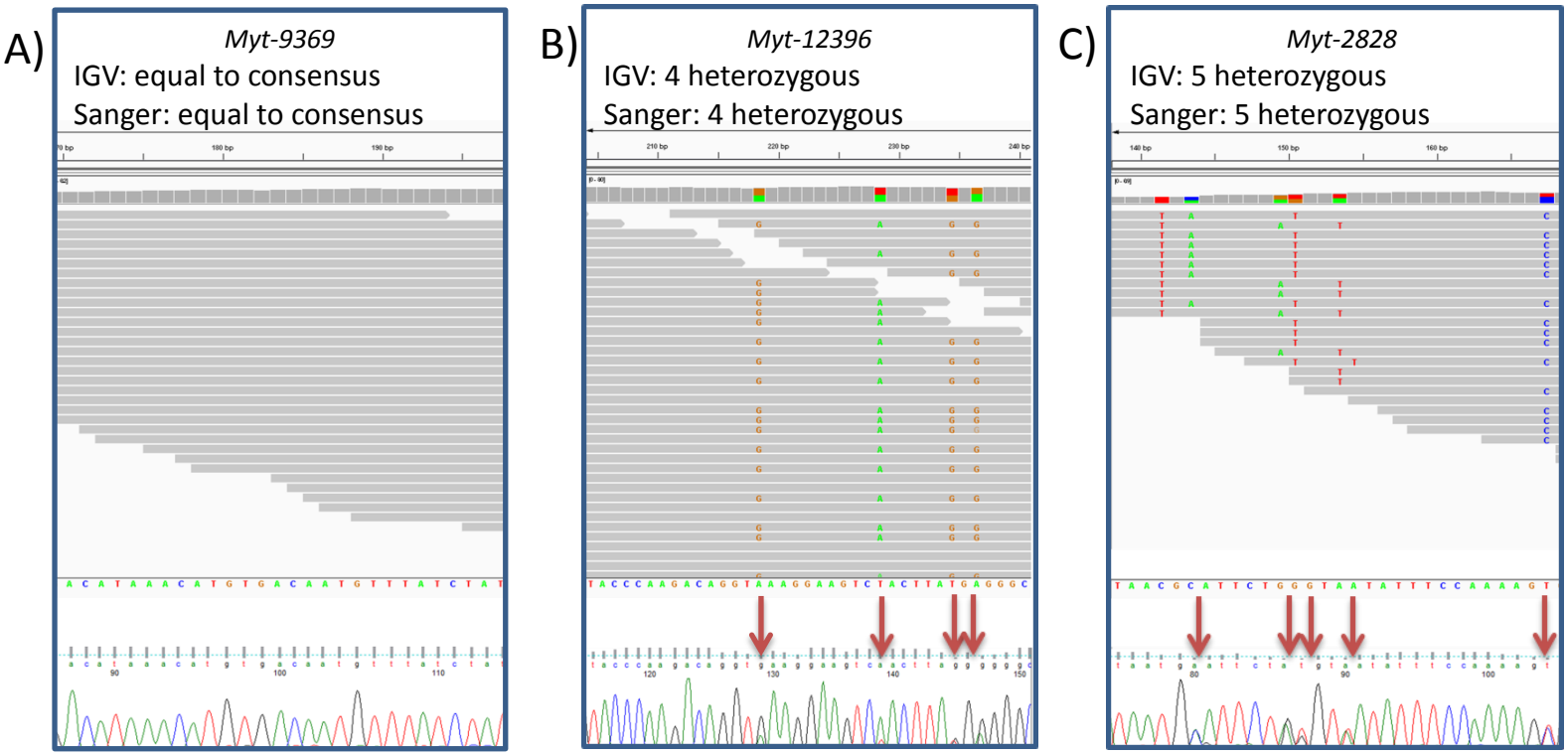


Figure 6. Comparison of Sanger and RAD sequences for RAD loci (A) consensus 9369 (B) consensus 12396 and (C) consensus 2828 in one individual homozygous for locus 9369 and heterozygous for 12396 and 2828. The upper sections are Integrative Genomic Viewer (IGV) visualisations of aligned reads at the RAD loci, where the grey represents consensus sequence and coloured bars show differences from the consensus. The lower sections show the chromatogram outputs of the Sanger sequencing with arrows highlighting the putative SNPs.

3.3.2.1.2 Mixture of mussel samples between families

Several measures were taken to ensure the integrity of mussel families was maintained. The production of families was performed under standard hatchery protocols to avoid cross-contamination of gametes: single-crosses were performed in independent, isolated tanks using filtered and UV radiated water. After fertilization, full-siblings were kept isolated throughout the extent of the experiment, first in tanks (in the hatchery) and then in independent pearl nets (in the estuary). Although the possibility of mixing of individuals between families was low, contamination is always a risk. To discard mixing of individuals as a potential explanation for observing (i) extensive heterozygote deficiency of discovered SNP markers and (ii) non-Mendelian inheritance of some SNP markers (*i.e*. non-parental alleles in the offspring), the validity of the pre-defined mussel family units was analysed using PCA. The rationale was that if the two families used for RAD-Seq were true biological units, this genetic relatedness should be evidenced by the genome-wide comparison of individuals' genotypes. Moreover, the two putative families were created by the single-cross of parent pairs that were translocated from distant natural populations; hence, due to geographical distance, most of the genetic variation is expected to be seen between, rather than within, family clusters.

The PCA was performed on the *population-dataset* to maximize the amount of genetic information utilised to establish individual's relatedness (or correlation). The first PC explains most of the variation in the dataset, 63%. The second and third PCs explain 11% and 7%, respectively. The graphing of the first and second PCs did not reveal a clear pattern of genetic structure (Fig. 7).

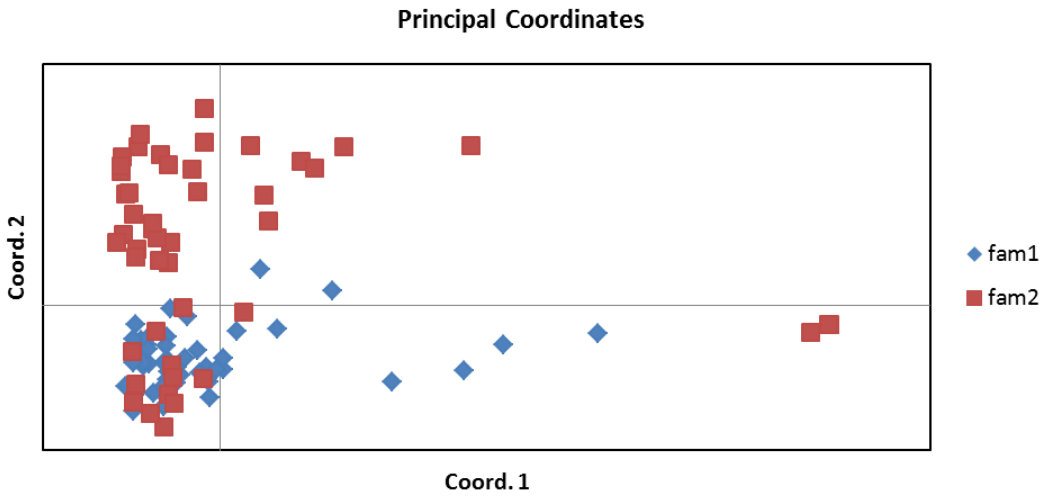


Figure 7. The PCA result of the two mussel families. The *x*-axis denotes the values of PC1 (63%), while *y*-axis denotes the value of PC2 (11%). Each dot in the figure represents an individual; each individual is coloured according to family provenance.

The graph of the first pair of PCs reflects little evident clustering according to family, with overlap of individuals from the two supposedly different families. If mussel families were true biological units, then this clustering might be indicative of recent or ancient admixture (interbreeding) between the populations. Another explanation for this clustering, given the assumption of reliability of family identity, is that the (supposedly) high *de novo* mutation rate that lead to extensive non-Mendelian inheritance was a true biological phenomenon, and that it is distorting the genetic relatedness (or correlation) within and among families. Alternatively, if the two putative families were not real biological families, then this clustering could have been indicative of cross-contamination of mussel samples, although it is difficult to deduce separate clusters from these results. Because of the multiple scenarios that could have originated the PCA results, conclusions on the potential mixture of mussel samples across families are unclear. However, it is worth mentioning at this stage that

segregation analyses of markers in experimental crosses of various bivalve species have frequently detected individuals that do not conform to Mendelian inheritance expectations at some loci (*i.e.* offspring with non-parental alleles) (MacAvoy et al., 2008, Del Rio-Portilla and Beaumont, 2000). Interestingly, the proposed argument to explain the presence of individuals harbouring odd alleles is cross-contamination; as such, these individuals are usually removed from further analysis. This suggests that biological contamination in hatchery based studies is of serious concern or that deviations from Mendelian expectations are an intrinsic factor of bivalve inheritance.

3.3.2.2 Biological explanations:

3.3.2.2.1 Intra-individual tissue variation in genotype: potential aneuploidy

Heterozygous offspring were observed to be far less frequent than expected in the offspring of heterozygous parents (*family-dataset*). One hypothesis that might explain this tendency towards homozygosis is variations in the copy number of the mussel genome. Two types of variation in copy number are known: polyploidy, which implies a numerical change in the whole set of chromosomes of an organism; and aneuploidy, which refers to a particular chromosome being under or over-represented. Typically, an extra set of chromosomes or chromosomal segment would lead to loci showing an excess of heterozygous genotypes, whereas a loss of a haploid copy or segment would lead to heterozygous deficiency. In bivalves, chromosomal loss has been observed in oysters (Thiriot-Quiévreux and Insua, 1992) and mussel natural populations (Dixon, 1982). Therefore, it is feasible that aneuploidy could be playing a role in the commonly observed heterozygote deficiencies.

As a first approximation to explore if the extensive non-Mendelian segregation ratio found in the mussel genome is originated by tissue-specific aneuploidy effects, the possibility of genetic variation between different tissues within individuals was analysed. Four whole-mussels (3 females and 1 male) were sampled, taking foot, gill (left and right), mantle and gonadic tissues for DNA extraction and genetic analysis. Each specific tissue was sequenced at three loci, *Myt-12396*, *Myt-2828*, and *Myt-9369*.

For each individual mussel, the DNA sequence obtained at each locus was compared (*i.e*. aligned to each other) across tissues. The analysis of four individuals revealed that at three sampled genomic regions (loci) no difference across tissue was detected. If aneuploidy was a source of heterozygote deficiencies, then a heterozygous SNP in a particular tissue would potentially be homozygous in another tissue, or vice versa; SNP genotypes are used as surrogate for the number of DNA strands contained in a particular tissue. Here, the agreement between tissue genotypes is understood as an indicative of absence of tissue-specific aneuploidy effects; however, the loss of chromosome(s) at the whole-individual level (rather that at a tissue-specific level) cannot be disregarded. An example of the alignment of locus *Myt-12396* amplified in different tissue is shown in Fig. 8.

Figure 8. A regional capture of the alignment of the locus *Myt-12396* across different tissues within an individual. Each nucleotide in the sequence has a specific colour. The letter 'S' (in grey) represents a heterozygous SNP. As can be observed, heterozygosity is maintained across the tissues.

3.3.2.2.2 Extended family testing for Mendelian Inheritance

The unusually high SNP frequency, occurrence of non-parental alleles and the frequent deviations from Mendelian segregation ratios were based on the analysis of family 1 only. Therefore, to test whether these phenomena applied more generally in the population, the study of the inheritance pattern of the discovered SNP markers was further extended to different mussel families.

The parents and 7 offspring of each family (6 families in total) were PCR-amplified with primer pairs *Myt-12396*, *Myt-2828*, and *Myt-9369*, Sanger Sequenced, and aligned for variant identification. The sequencing of 3 mothers failed with all three primer pairs tested, leading to a reduction of the inheritance analysis to three, instead of six, family structures. Apart from the high SNP frequency observed within and across families, several unexpected alleles were observed in the offspring. The distribution of the unexpected alleles within each family is apparently non-random; an unexpected SNP in one of the offspring is likely to appear in the same position in a sibling mussel (Fig. 9). If the family units were reliable, this observed pattern of segregation (inheritance)

supports a high rate of *de novo* mutations, at least for the three sequenced loci analysed. This tendency of mutations (or non-parental SNP alleles) to appear in the same genomic region among sibling mussels may be explained by a point mutation appearing early in the process of gamete formation. Additional information from other genomic regions is required to examine the extent to which non-parental alleles appear in the offspring.

The additional experiments from Stage 2 indicate that the high SNP frequency found in the mussel genome by RAD-Seq reflects true mussel genetic variation (validation by Sanger sequencing). Regarding the extensive heterozygote deficiencies and frequency of non-parental alleles appearing in the offspring, our results are not conclusive. This is mainly because although there is no reason to believe that mixture of mussel individuals occurred, the possibility cannot be discarded. The nature of the suggestive findings from the current study (non-parental alleles, unusual segregation ratios) precludes the conclusive test of family assignment by standard genetic analysis. If mussel family units are not true full-sibling families, then downstream analysis examining inheritance of SNP markers would alter potential conclusions. However, numerous population studies of marine bivalves have reported deficiencies of heterozygotes relative to expectations (*e.g*. under Hardy-Weinberg equilibrium). Moreover, studies on marker inheritance in oyster and mussel families have also observed non-parental alleles in the offspring, but however these invoke mixing of individuals during hatchery operations as the most plausible explanation. The fact that heterozygote deficiencies and non-parental offspring alleles have been reported previously, in a smaller scale (few molecular markers) for other bivalve species, is evidence that supports that our observations are probably not caused by mixing of individuals. If mussel families were true biological units, then one possibility that could explain extensive heterozygote deficiencies and non-Mendelian inheritance of some

SNP markers are *de novo* constitutional mutations (*i.e.* mutations present in every cell) occurring in the gonads during gamete formation. Due to the extensive cell division that the bivalve germ line epithelium undergoes to achieve the thousands of gametes released every reproductive event, a high level of *de novo* mutations are expected.

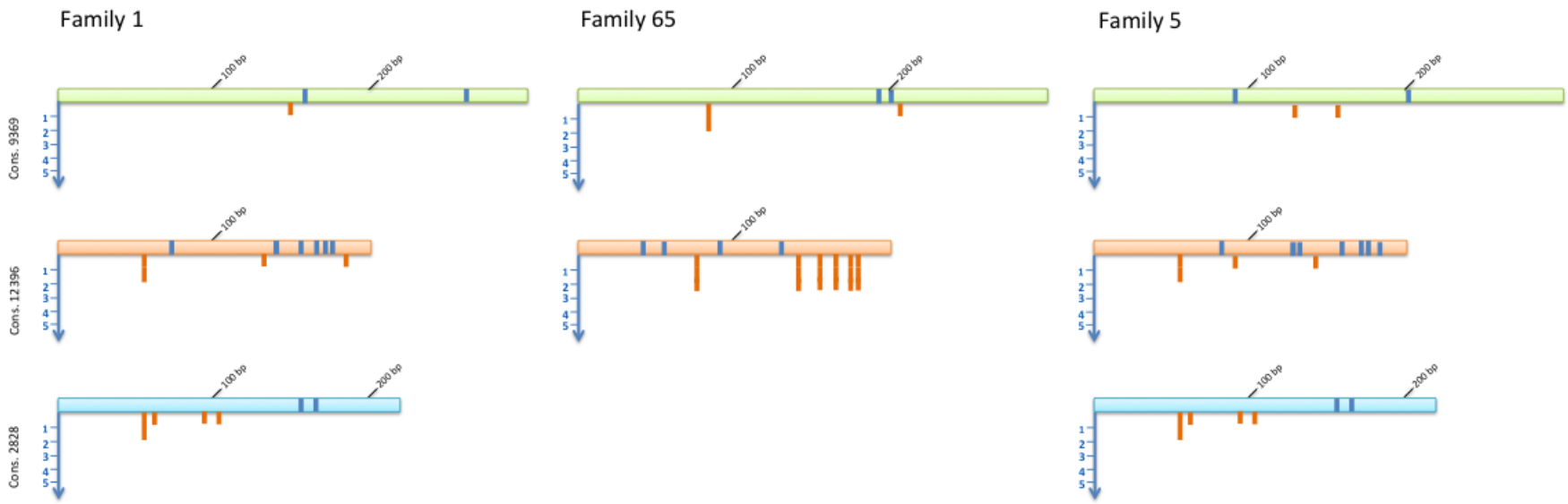


Figure 9. Representation of the genetic variation within three genomic regions (RAD consensus loci 9369, 12396, and 2828) across three mussel families (parents + 7 offspring). The coloured horizontal bars represent the consensus sequences at the loci and the blue boxes indicate the position of a variable position in one or both parents (*i.e*. indicating a SNP). Orange boxes (under each sequence) indicate variable positions in the offspring where both parents are fixed for the consensus allele, and the Y axes are counts of offspring showing a non-parental allele at that position

3.4 Conclusions

RAD-Seq was applied in two mussel families for high throughput SNP discovery and genotyping: 4,537 SNPs were found segregating in one family when constraint on parental heterozygosity was imposed; 65,553 SNPs were detected across all samples (both families) without any constraint on parent genotype. The SNP frequency (approx. 1 SNP each 29 bp in the *population* dataset) is one of the highest observed among metazoan phyla. The segregation (inheritance) analysis within a mussel family revealed that most SNP markers (72%) showed a deviation from expected genotype ratios under Mendelian inheritance, with a tendency towards a deficiency of heterozygous genotypes. Further, a high number of novel (*de novo*) SNP alleles were observed in the mussel offspring without being present in their parents. The distribution of *de novo* SNPs is apparently non-random; a non-parental SNP allele in one of the offspring is likely to appear in the same position in a sibling mussel.

Considerable effort was placed into exploring the validity of these findings and searching for interpretations. Independent confirmation of the results obtained in the RAD-Seq (using the Illumina HiSeq 2000) by Sanger technology indicates that the RAD-Seq data and the SNP calling protocol were reliable. A mixture of individuals between families cannot be discarded by the results of the PCA. However, the security measures taken throughout the experiment to avoid cross-contamination suggests that the chance of a mixture was minimal. The comparison between genotypes across tissues indicates that aneuploidy (*i.e.* chromosomal loss) is not a plausible explanation for observing the deficiency of heterozygous genotypes. Finally, the unusual segregation ratios and frequent non-parental alleles were confirmed in several additional families.

CHAPTER 4

Heritability of growth-related traits in the Chilean mussel

4.1 Introduction

The aquaculture of the Chilean mussel relies heavily on obtaining mussel seeds from the natural environment; therefore the availability of seeds is determined by the reproductive capacity of wild mussel populations, which is unreliable on a year-to-year basis. Consequently, large-scale mussel seed supply through hatchery technology is an important short-term goal for the mussel industry, concomitantly offering the opportunity for developing selective breeding programmes for the Chilean mussel. Although mussel hatcheries have been established in New Zealand, Spain and Canada [in Uriarte (2008)], data on selection response for growth rate (usually the primary breeding goal) are not available. However, mussels' exhibit high heritabilities for growth rate [h^2=0.9; Strömgren and Nielsen (1989)] and shell length [h^2 from 0.4 to 0.9; Toro and Paredes (1996)]. Therefore, a breeding program for growth is expected to bring significant progress to the mytiliculture industry.

However, in Chapter 3 the analysis of the mussel genome suggested that much genetic variation is not inherited from parents to offspring in accordance with Mendelian expectations, and apparently novel genetic variants appear frequently in the offspring. If this *de novo* genetic variation is dispersed extensively throughout the genome – a fact that cannot fully be explored with the reduced representation genome sequencing method used in this thesis (RAD-Seq) – an effect on the heritability of phenotypes may be expected. In principle, if genes are influencing a phenotype, then biological relatives

should, on average, resemble one another more than unrelated individuals do. This genetic principle is the basis of traditional heritability estimations: by analysing the phenotypic similarity between relatives (*e.g*. correlation between parents and offspring or full-siblings), the variance of an observed phenotype (V_P) can be partitioned in heritable, additive genetic (V_A) and non-heritable, environmental (V_E) variation. These variance components are then used in the estimation of trait heritability, which is defined as the proportion of the total observed variation in a phenotype that is attributed to additive (heritable) genetic variation in a population.

i.e., $h^2 = V_A/V_P$

V_A = additive genetic variance

V_P = phenotypic variance

Assumptions relating to the nature of between-family phenotypic variation are not always straightforward. For example, if high levels of sporadic non-inherited variation are being randomly created in each generation, the overall individual expression of genes (or gene variants) may be compromised, which in turn may have a significant influence on the phenotype of an individual. Altogether these individual-level effects can lead to a more generalised one, primarily, an increase in the phenotypic variance of a population. If the phenotypic variance of a trait is increased by means of *de novo* genetic variation, a reduction in phenotypic resemblance between relatives can be expected, which will reduce the magnitude of additive gene effects, and therefore heritability estimates are likely to decrease.

In Chapter 3, a molecular genetic approach was used to discover and describe a phenomenon in the mussel genome, namely extensive heterozygote deficiencies and high levels of non-parental (*de novo*) alleles. In the face of a bivalve species exhibiting

high levels of *de novo* variation, I wanted to investigate further whether this phenomenon had a population-level effect, particularly on the additive genetic variance components of phenotypes. Therefore, the aim of this study was to quantify the relative importance of additive genetic components on two growth-related traits (shell length (SL) and body weight (WT)) by estimating their heritabilities at two time points.

4.2 Material and Methods

4.2.1 F1 Resource Families and Phenotypic records

The production of F1 resource mussel families was described in Chapter 2. From the 150 mussel families created by single pair crossing, 31 families from CH, 21 families from YA, and 27 families from PA were used in the present study. Each family consisted of the sire, the dam, and ~280 offspring (total n~22,278).

Phenotypic records of individual shell length (SL) and wet body weight (BW) were recorded at 10 and 16 months of age. The SL was measured using an electronic vernier calliper (0.01 mm accuracy), and BW was measured using an electronic balance (0.1 g accuracy). The number of offspring records was not constant between both time points, mainly because of individual mortality. On average, 280 individual mussels per family were measured for SL and WT at 10 months of age, while an average of 198 individuals per family remained at 16 months and were measured for SL and BW. Whilst individuals were recorded at the family level, individual identities were not kept, hence it was not possible to determine which individuals remained nor was it possible to estimate correlations across time.

4.2.2 Data analysis

The linear relationship between SL and BW was analysed by estimating Pearson's correlation coefficient. In addition, descriptive statistics of both traits were performed using SAS software. Since SL measurements were not normally distributed in the population, the data was log transformed. The heritability values of each phenotype ($h^2=V_A/V_P$, where V_A and V_P are the additive genetic and the phenotypic variance, respectively) at two time points were calculated from the partitioning of variance components obtained by fitting a single-trait animal model using the ASReml package (Gilmour et al., 2006). The model was as follows:

$Y_i = \mu + a_i + e_i$

Y_i is the SL or BW of the i^{th} individual

μ is the overall mean of the trait

a_i is the additive effect of the i^{th} animal

e_i is the residual term

4.3 Results and Discussion

The heritability of BW and SL was estimated in 79 full-sibling mussel families at two time points with the aim of estimating the additive genetic component of both traits. The correlation between the traits was low ($r=0.23$, $P<0.01$), which is partially reflected by the difference of their heritability estimates. For SL, the heritability estimates were low and non-significant 10 and 16 months (~0.01). However, the heritability of BW was highly significant and appeared to increase with age. BW heritabilities were 0.11 and 0.28 at 10 and 16 months of age, respectively (Table 1). Heritability of a trait can

change with age for example due to changes in the environmental condition (Falconer and Mackay, 1996). Moreover, in this particular study trait heritabilities were estimated in a mussel population grown in a natural environment, therefore, high mortalities (near 70%) were observed. This elimination of individuals from the population may have affected trait heritability estimations, particularly if the loss was genotype-dependant rather than random.

Table 1 Mean and heritability (h^2) with respective standard deviations (SD) for Body Weight (BW) and Shell Length (SL) at two ages

Age (months)	Trait	Mean± SD	h^2 ± SD
10	BW	1.35 ± 0.4	0.11 ± 0. 02
	SL	23.63 ± 17.8	0.01 ± 0. 00
16	BW	8.58 ± 2.5	0.28 ± 0. 04
	SL	46. 02 ± 42.7	0.01 ± 0. 00

The heritability of BW was low and estimated to be 0.11 and 0.28 at 10 and 16 months, respectively. Whereas the heritability of SL was near zero at both time points. In general, heritability estimates of economically important traits in mussels are highly variable. For example, in the Australian mussel (n= 1,538) the heritability of body weight, meat yield and shell shape in the Australian mussel were low (0.051, 0.049 and 0.085, respectively) (Nguyen et al., 2011). Similarly, Alcapan and collaborators (2007) reported that heritability for body weight in the Chilean mussel was low but increased with age from 0.001 to 0.150 at 12 and 22 months of age, respectively. In this same

species, in contrast, Toro and co-workers (2004) reported that body weight and shell height heritabilities were higher $h^2 = 0.35$ and $h^2 = 0.32$, respectively. Since we found that BW was partially affected by heritable variation in genotypes (*i.e*. by additive gene effects), then selection of this particular mussel population for increased BW will lead to genetic progress in the next generation. This is despite the fact that the trait heritabilities were estimated in the same population from which we sampled and sequenced the families with extensive *de novo* genetic variation and non-Mendelian inheritance ratios. Therefore, despite the unusual inheritance features of the mussel genome, exploitable additive genetic variation exists at a population level. However, this raises the question of whether the magnitude of heritability estimates is influenced by *de novo* variation, and therefore the real improvement that can be achieved by the application of selection programmes. If this *de novo* variation have a marginal effect on the biology of mussels, *e.g*. if they are found mainly in non-functional genomic regions, then genetic improvement could, in principle, be straightforward. However, if SL, BW and other phenotypes of potential commercial interest are potentially being disturbed by the *de novo* variation, then their improvement through selection is unclear.

CHAPTER 5

General Discussion

5.1 Introduction

Genetics and genomics techniques have immense potential for enhancing aquaculture production through selective breeding programs, including the incorporation of MAS. Key components that are required for an efficient MAS system include (i) established marker-trait associations for traits of interest, (ii) suitable genetic markers and their characterization and (iii) high-throughput genotyping systems, among others. The general aim of this Master of Philosophy thesis was to develop molecular resources that may prove useful for MAS in two aquacultural species of economic importance, the Atlantic salmon and the Chilean mussel. Since both species are in different stages of development of genomic research, specific aims were proposed according to the genomic resources available and the needs of each industry.

5.1.1 Atlantic Salmon

For the Atlantic salmon a candidate gene association study was performed. Two gene paralogues (MSTN-1a and MSTN-1b), which are suggested to play a role in muscle regulation in teleost fish, were screened for gene variants to directly test their association with several traits measured at harvest in a large commercial population (n=4,800) (Chapter 2). Three SNPs were detected on the MSTN-1b gene. One SNP, which was located within the 5' flanking region of the gene (g.1086C>T), showed a significant association with all the body-weight traits (Harvest Weight, Gutted Weight, Deheaded Weight and Fillet Weight), all of which show h^2 above 0.5 and a high positive

phenotypic correlation (r>0.97). The alleles in the g.1086C>T locus acted in an additive matter, with a change from a CC->TT genotype associated with an increase of 70 to 100g depending on the trait. This research provides an example of the development of a genetic resource (*i.e.* the g.1086C>T locus) that can be used in selection schemes to improve growth and fillet related traits in commercial populations of Atlantic salmon via the use of MAS.

Candidate gene association studies are relatively uncommon in salmon research which may be related to a lack of knowledge of the biological regulation of economically important phenotypes. However, they have the advantage – compared to other approaches such as genome-wide association studies (GWAS) – of being capable of detecting markers with smaller associated effects due to the lack of requirement for stringent multiple-testing corrections. The MSTN-1b polymorphism g.1086C>T was found to explain a small proportion of the variance in salmon body-weight traits in an association study with a high statistical power (n=4,800). Failure to locate SNPs that explained a higher percentage of the trait(s) variance(s) could be related with the candidate gene itself, that is, the myostatin gene. In mammals MSTN is known as a strong negative regulator of muscle mass (Lee and McPherron, 1999). However, in teleost fish up to four MSTN gene paralogues have been detected. In addition, the expression pattern of MSTN paralogues and their regulation are different from those reported in mammals [reviewed in Gabillard and co-workers (2013)]. Altogether, this evidence suggests that probably fish MSTN paralogues do not exert as strong a regulatory role as seen in their mammalian counterpart. In the context of our study, the small variance explained by g.1086C>T might reflect that the candidate gene MSTN-1b plays a more general biological function (*e.g*. as an inhibitor of cell proliferation and cell growth), and is not specialized, like in mammals, into a strong muscle regulator. Alternatively, it is also possible that MSTN-1b effectively has a significant negative

regulatory role in muscle development, but however, because of selection constraints on fish's hydrodynamic shape, finding a 'double-muscled' phenotype (and therefore SNPs with bigger effects), is not biologically feasible. If this is the case, then fish populations will encode more stable MSTN-1b products, with underlying gene variants exerting more conservative effects on muscle-related phenotypes. Whether the small variance explained by g.1086C>T is, to some extent, indicative of the magnitude of MSTN-1b regulatory role will remain unclear until additional functional studies become available.

Despite the small percentage of variation explained, the results presented herein add to the evidence that suggests SsMSTN-1b, and potentially the orthologous gene in other teleost species, plays a role (of unknown magnitude) on muscle development in fish. Since g.1086C>T is significantly associated with growth-related traits, then it is likely to be useful in a panel of DNA tests to improve overall performance of Atlantic salmon. The combined effect of further QTL–mapping and candidate gene studies assessing the association between genotypes and growth traits may further unravel genes of larger effect and lead to an improved understanding of the regulation of muscle growth in fish.

5.1.2 Chilean mussel

Basic genetic and genomic resources for most aquacultured species are still lacking. Furthermore, species do not always fit *a priori* expectations based on current knowledge and sometimes show unusual patterns of inheritance and population genetic parameters. For these species, although the development of molecular resources is necessary, most important is achieving an understanding of the mechanisms underlying these observations.

This is the case for marine bivalves, a group of species with scarce genetic and genomic resources that exhibit a particularly complex genetic pattern, *i.e.* heterozygote deficiencies. This potentially impacts on the application of markers to breeding programs for family assignment or marker-based selection methods.

The overall aim for the Chilean mussel experiments described in this thesis was to assess the possibility of selective breeding for growth-related traits by assessing their heritability (Chapter 4) and by discovering and characterising the inheritance of genetic markers in mussel families (Chapter 3).

In Chapter 4 the heritability of two growth-related traits, shell length (SL) and body weight (BW), was estimated at two ages (10 and 16 months) in 79 mussel families (total n~22,278). The heritability of BW was low and estimated to be 0.11 and 0.28 at 10 and 16 months, respectively. On the other hand, the heritability of SL was near zero at both time points. Although the heritability of SL indicates that most of the variance in this trait is environmental, the heritability results for BW indicate that it is possible to increase weight gain in mussel populations by selective breeding. This information is useful considering that the mytiliculture industry in Chile is now moving towards hatchery technologies for ensuring year-round seed supply. By controlling the breeding stage of mussels, exploitation of the genetic variation underlying traits of economic importance will become feasible, thereby leading to a significant improvement of productivity by means of selective breeding.

MAS can improve the efficiency of selective breeding for traits with low heritability (*e.g*. as BW) by controlling the selection time, selection pressure and accuracy. The first step to apply MAS in a mussel breeding program is to develop and characterise high-density SNP markers. In Chapter 3 two Chilean mussel families were used for genome-wide SNP discovery and high-throughput genotyping using RAD-Seq. Originally, 4,537 SNPs

were identified in the assembled RAD-loci by applying a requirement on parental genotypes - SNPs were called only if one or both parents were heterozygous for the locus. The characterisation of these SNP markers led to three main findings. First, the mussel genome is highly polymorphic. In fact, the Chilean mussel SNP frequency is one of the highest described to date in a metazoan species - 1 SNP each 29 bp. Second, the majority of discovered SNP markers (72%) deviate from Mendelian expectations. In particular, offspring showed a lower frequency of heterozygous genotypes than expected from parental genotypes. And third, offspring exhibit novel SNPs compared to their supposed parents.

These observations (high SNP frequency, heterozygote deficiency and offspring showing *de novo* alleles) have been reported previously in literature on bivalve genetics; however not previously at the genome-wide scale observed in this thesis. The wide amount of genetic data available on mussel families provided the opportunity to develop further experiments to examine whether the explanations for these observations were of technical or biological nature. The main results from these additional experiments indicated that: (i) the high SNP frequency detected by NGS-based RAD-Seq reflects true mussel genetic variation, as confirmed by Sanger sequencing; (ii) mixing of individuals between families cannot be disregarded as a possible explanation for observing both heterozygote deficiency and novel alleles in the offspring; (iii) tissue-specific chromosomal loss (aneuploidy) apparently cannot explain heterozygote deficiency because heterozygous SNP loci were consistently found across several tissue of individual mussels; and (iv) the distribution of *de novo* SNPs among full-siblings is apparently non-random, new polymorphisms tend to appear at the same nucleotide position across offspring.

The interpretation of these results is complex, mainly because the potential mixing of individuals between mussel families could not be discarded by genetic analysis. However, several security measures were taken during the experiment to avoid cross-contamination of individuals; therefore, the possibility of a mixture between mussel families was believed to be minimal. Moreover, if mixing really occurred, it would have had to have been significant and widespread in order to explain the results. Therefore, I consider it to be an unlikely explanation for the results. Since other family-based studies on bivalves have observed similar genetic features, it is reasonable to believe that our putative mussel families are true biological units, and therefore the results presented herein effectively reflect features of genetic transmission (inheritance) in mussels. The fact that significant h^2 estimates were obtained for body weight supports the conclusion that families had been accurately identified and most probably remained unmixed in our experiments.

The analysis of the segregation pattern of SNP markers led to the identification of several *de novo* polymorphisms in the mussel progeny. In principle, if a polymorphism has not been inherited from either parent, then it probably was exogenously or endogenously induced. A possible mechanism for explaining polymorphisms being induced in a genome are *de novo* mutations arising during gametogenesis. *De novo* mutations might also be related by causality with the observed high SNP frequencies. Gametogenesis is the biological process by which precursor cells in the gonad undergo cell division and differentiation to form mature haploid gametes. During this process – and mediated by cell division – the gamete genome acquires mutations due to spontaneous errors in DNA replication and repair. The number of germ line mutations not only depends on the number of cell divisions, but also depends on the reproduction mode of the species. Mussels (and bivalves in general) sexually reproduce by spawning thousands of gametes into the water column each year. It follows that the number of

cell divisions in germ line does not show a marked asymmetry between males and females as seen, for example, in humans and chimpanzees, for which whole-genome sequencing has reported a six-fold difference in mutation rate between male and female germlines [in Conrad et al. (2011)]. As in mussels both sexes are equally producing large numbers of gametes, with an underlying large number of cell divisions occurring in both germ lines, then high levels of inherited and novel genetic variation are expected in each cohort. This expectation is concordant with what was observed in Chapter 3 – high levels of genetic polymorphisms, both inherited and created (novel), were detected throughout the Chilean mussel genome.

Chapter 3 also reported that ~70 % of the discovered SNP markers showed a deficiency of heterozygous individuals compared to Mendelian expectations. This unusual phenomenon has been a long-standing observation in population and family-based studies on bivalve species. Several non-mutually exclusive hypotheses have been proposed to account for bivalve's heterozygote deficiencies. Yet two are the most invoked - selection of heterozygote genotypes at early larval stages and null-alleles leading to the misscoring of heterozygotes as homozygote genotypes. Interestingly, the visual inspection of the aligned RAD-loci within each mussel family lead to a curious observation regarding the genetic transmission of markers. Because mussels exhibit high levels of genetic polymorphism, within a single RAD-loci (sequence near a SbfI restriction site) several SNPs were found, thereby allowing the phasing of each individual's constituent haplotypes. The analysis of haplotype inheritance within the mussel family revealed that the pattern of genetic diversity in the offspring along with their heterozygote deficiencies fit to a mechanism in which the haplotype of one parent is not seen in some progeny (Offspring 4, Fig. 1). In other cases, (*e.g*. Offspring 6) completely new haplotypes are observed in the offspring that do not exist in either parent.

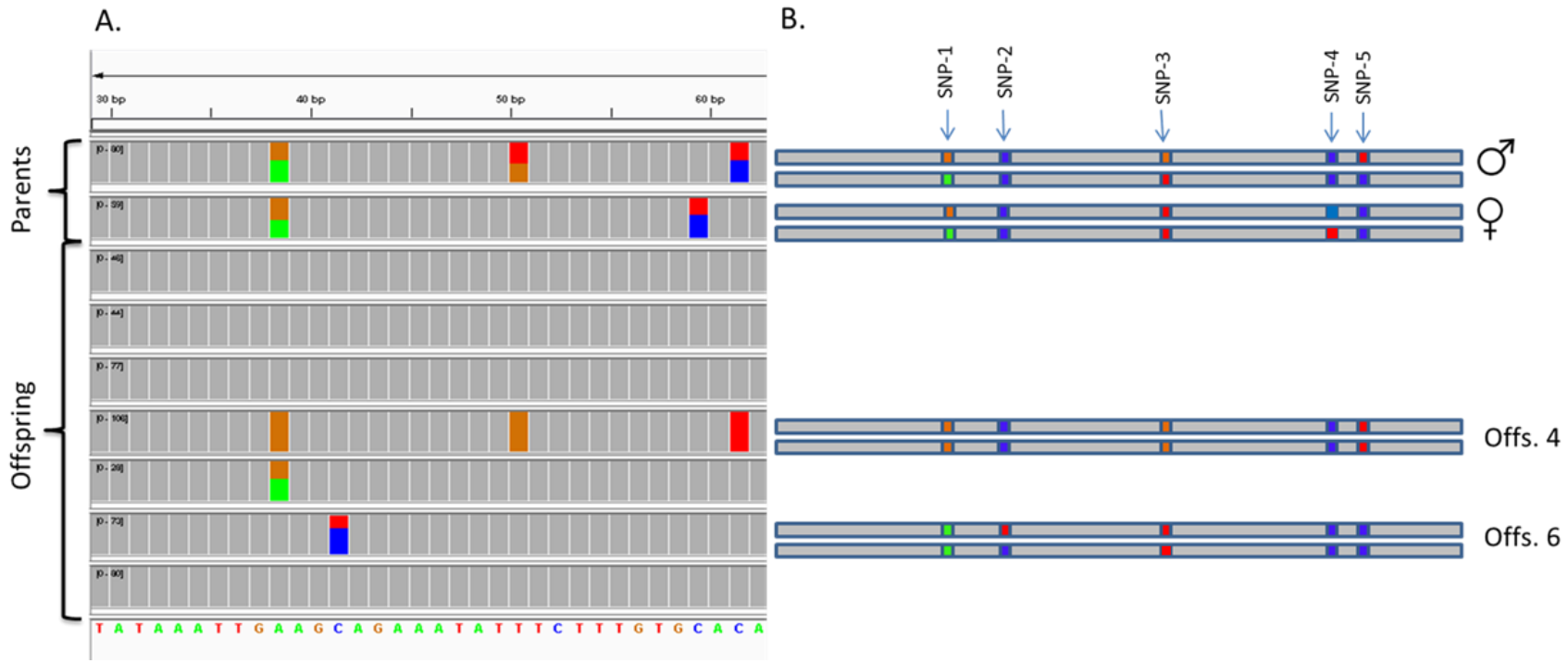


Figure 1. SNP inheritance in a mussel family. A. Integrative Genomic Viewer (IGV) regional capture showing the alignments for a RAD-locus in mussel parents and seven offspring. Coloured blocks represent a SNP polymorphism, each colour represents a specific nucleotide: A (green), G (mustard), T (red) and C (blue). When a block is not coloured it means that the nucleotide position coincides with the consensus sequence for that RAD-locus, the consensus sequence is shown below the alignment. B. Haplotype reconstruction for the mussel parents and two offspring. All genetic variation within the region is coloured in order to enable the visualization of the inheritance pattern

This observation is in striking contrast to typical Mendelian inheritance in diploid organisms, for which haplotypes are being inherited one from each parent, and therefore, under current assumed knowledge that inheritance in bivalves is Mendelian, highly improbable. However, it must be noted that bivalves are organisms that do not always conform to the traditional expectations of diploid animals. Notably, bivalves are the only group of organisms in which mitochondrial DNA (mtDNA) shows a double uniparental inheritance (DUI) (Zouros et al., 1994). Most metazoan mtDNA is known to be usually transmitted by strict maternal inheritance (Avise, 1991); however, over 36 bivalve species deviate from this rule (Theologidis et al., 2008) and exhibit two mitochondrial genomes coexisting in a stable state. One mitochondrial genome, known as maternal (symbolized as F), is transmitted from mothers to both female and male progeny. The other mitochondrial DNA is paternal (symbolized as M), and is transmitted from males to their male offspring. The F mtDNA of a male parent cannot be transmitted to its progeny. Moreover, the distribution of both mtDNAs in adult mussels is tissue specific. In females, the F mtDNA is the prevalent genome in all tissues, although marginal amounts of M mtDNA can also be found. In males the F mtDNA is also predominant in various somatic tissues. However, the gonad is dominated by the M mitochondrial genome (Garrido-Ramos et al., 1998). This unique and fascinating mechanism of organelle transmission, especially in the face of the overall results presented herein, at least raises doubts that nuclear inheritance in mussels is strictly Mendelian.

The purpose of Chapter 3 was to develop genomic resources for the Chilean mussel, envisioning their application in future breeding programs. Although the aim was achieved, the characterisation of the discovered SNPs enabled the realization that more fundamental questions have to be addressed before the promise of MAS can transform the mytiliculture industry – mainly, how genetic variation is being created and

transmitted in mussels. It is noteworthy to mention that this understanding is not only relevant for breeding purposes, it is also relevant for phylogenetic and population genetic studies in bivalves. For example, in phylogenetic and clustering analysis if *de novo* mutations are significantly contributing to mussels high genetic polymorphism – which has yet to be proven – then clades that are not the evolutionary closest relatives will cluster together due to misinterpretation of shared homoplasies. This artefact is known as Long Branch attraction (LBA) and can seriously mislead topological interpretation. Since population genetics and phylogenetic inference methods are used for the delimiting species and identifying population stock structures – knowledge that is applied to conservation and resource management – then unusual sources of genetic inheritance and/or variation has to be taken in consideration in routine genetic analysis of bivalve species.

In this research the genome-wide segregation pattern of a bivalve species was studied at an unprecedented scale, revealing that unusual genetic patterns described previously for bivalves are in fact a strong, widespread genomic phenomenon. This opens new avenues of investigation at the time that challenges our understanding of one of the most fundamental processes of life, namely how genetic information is being transmitted in marine bivalves.

Bibliography

ALCAPAN, A., NESPOLO, R. & TORO, J. 2007. Heriability of body size in the Chilean blue mussel (Mytilus chilensis Hupe 1984): effects of environment and ageing. *Aquaculture Research,* 38**,** 313-320.

ALLENDORF, F. & THORGAARD, G. 1984. Tetraploidy and the Evolution of Salmonid Fishes. *In:* TURNER, B. (ed.) *Evolutionary Genetics of Fishes.* Springer US.

ALTSCHUL, S., MADDEN, T., SCHÄFFER, A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res,* 25**,** 3389-3402.

AMORES, A., FORCE, A., YAN, Y.-L., JOLY, L., AMEMIYA, C., FRITZ, A., HO, R. K., LANGELAND, J., PRINCE, V., WANG, Y.-L., WESTERFIELD, M., EKKER, M. & POSTLETHWAIT, J. H. 1998. Zebrafish hox Clusters and Vertebrate Genome Evolution. *Science,* 282**,** 1711-1714.

ANDREASSEN, R., LUNNER, S. & HØYHEIM, B. 2009. Characterization of full-length sequenced cDNA inserts (FLIcs) from Atlantic salmon (Salmo salar). *BMC Genomics,* 10**,** 502.

ANISIMOVA, A. 2007. Genome sizes of some Bivalvia species of the Peter the Great Bay of the Sea of Japan. *Comparative Cytogenetics,* 1**,** 63-69.

ASCHE, F. & BJORNDAL, T. 2011. *The economics of Salmon Aquaculture*, John Wiley & Sons.

ASCHE, F., H. ROLL, K., N. SANDVOLD, H., SØRVIG, A. & ZHANG, D. 2013. Salmon aquaculture: Larger companies and increased production. *Aquaculture Economics & Management,* 17**,** 322-339.

AVISE, J. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Review of Genetics,* 25**,** 45-69.

BAGNARA, M. & MALTRAIN, G. 2008. Descripcion del sector miticultor en la region de Los Lagos, Chile: evolucion y proyecciones. *In:* LOVATELLI, A., FARIAS, A. & URIARTE, I. (eds.) *Estado actual del cultivo y manejo de moluscos bivalvos y su proyeccion futura.* Rome: FAO.

BAIRD, N., ETTER, P., ATWOOD, T., CURREY, M., SHIVER, A., LEWIS, Z., SELKER, E., CRESKO, W. & JOHNSON, E. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One,* 3.

BARANSKI, M., MOEN, T. & VÅGE, D. 2010. Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (Salmo salar). *Genet Sel Evol,* 42**,** 17.

BARRETT, J. C., FRY, B., MALLER, J. & DALY, M. J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* 21**,** 263-5.

BAXTER, S., DAVEY, J., JOHNSTON, J., SHELTON, A., HECKEL, D., JIGGINS, C. & BLAXTER, M. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One,* 6.

BEAUMONT, A. R. 1991. Genetic studies of laboratory reared mussels, Mytilus edulis: heterozygote deficiencies, heterozygosity and growth. *Biological Journal of the Linnean Society,* 44**,** 273-285.

BEAUMONT, A. R. & BEVERIDGE, C. M. 1984. Electrophoretic survey of genetic variation in Pecten maximus, Chlamys opercularis, C. varia and C. distorta from the Irish Sea. *Marine Biology,* 81**,** 299-306.

BENSON, D. A., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & WHEELER, D. L. 2007. GenBank. *Nucleic Acids Res,* 35**,** D32-D37.

BORSA, P., ZAINURI, M. & DELAY, B. 1991. Heterozygote deficiency and population structure in the bivalve Ruditapes decussatus. *Heredity,* 66**,** 1-8.

BOSTOCK, J. 2010. The application of science and technology development in shaping current and future aquaculture production systems. *The Journal of Agricultural Science,* 149 133-141.

BOULDING, E., CULLING, M., GLEBE, B., BERG, P., LIEN, S. & MOEN, T. 2008. Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. *Heredity (Edinb),* 101**,** 381-391.

BROOKFIELD, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology,* 5**,** 453-455.

BRUGERE, C. & RIDLER, N. 2004. *Global aquaculture outlook in the next decades: an analysis of national aquaculture production forecast to 2030,* Rome, FAO Fisheries Circular.

CALEY, M., CARR, M., HIXON, M., HUGHES, T., JONES, G. & MENGE, B. 1996. Recruitment and the local dynamics of open marine populations. *Annual Review of Ecology and Systematics,* 27**,** 477–500.

CAMPBELL, C. & EICHLER, E. 2013. Properties and rates of germline mutations in humans. *Trends in genetics : TIG,* 29**,** 575-584.

CÁRDENAS, L., SÁNCHEZ, R., GOMEZ, D., FUENZALIDA, G., GALLARDO-ESCÁRATE, C. & TANGUY, A. 2011. Transcriptome analysis in Concholepas concholepas (Gastropoda, Muricidae): mining and characterization of new genomic and molecular markers. *Marine Genomics,* 4**,** 197-205.

CARPIO, Y., ACOSTA, J., MORALES, R., SANTISTEBAN, Y., SANCHÉZ, A. & ESTRADA, M. 2009. Regulation of body mass growth through activin type IIB receptor in teleost fish. *General and comparative endocrinology,* 160**,** 158-167.

CATCHEN, J., AMORES, A., HOHENLOHE, P., CRESKO, W. & POSTLETHWAIT, J. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda, Md.),* 1**,** 171-182.

CATCHEN, J., HOHENLOHE, P., BASSHAM, S., AMORES, A. & CRESKO, W. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol,* 22**,** 3124-3140.

CHARLESWORTH, B. & LANGLEY, C. H. 1989. The population genetics of Drosophila transposable elements. *Annual Review Of Genetics,* 23**,** 251-287.

CHING, A., CALDWELL, K., JUNG, M., DOLAN, M., SMITH, O., TINGEY, S., MORGANTE, M. & RAFALSKI, A. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet,* 3**,** 19.

CLABBY, C., GOSWAMI, U., FLAVIN, F., WILKINS, N. P., HOUGHTON, J. A. & POWELL, R. 1996. Cloning, characterization and chromosomal location of a satellite DNA from the Pacific oyster, Crassostrea gigas. *Gene,* 168**,** 205-209.

CLOP, A., MARCQ, F., TAKEDA, H., PIROTTIN, D., TORDOIR, X., BIBÉ, B., BOUIX, J., CAIMENT, F., ELSEN, J.-M., EYCHENNE, F., LARZUL, C., LAVILLE, E., MEISH, F., MILENKOVIC, D., TOBIN, J., CHARLIER, C. & GEORGES, M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet,* 38**,** 813-818.

CONRAD, D., KEEBLER, J., DEPRISTO, M., LINDSAY, S., ZHANG, Y., CASALS, F., IDAGHDOUR, Y., HARTL, C., TORROJA, C., GARIMELLA, K., ZILVERSMIT, M., CARTWRIGHT, R., ROULEAU, G., DALY, M., STONE, E., HURLES, M., AWADALLA, P. & GENOMES, P. 2011. Variation in genome-wide mutation rates within and between human families. *Nature Genetics,* 43**,** 712-714.

CORTE-REAL, H. B. S. M., HOLLAND, P. W. H. & DIXON, D. R. 1994. Inheritance of a nuclear DNA polymorphism assayed in single bivalve larvae. *Marine Biology,* 120**,** 415-420.

CRENSHAW JR, J. W., HEFFERNAN, P. B. & WALKER, R. L. 1996. Effect of growout density on heritability of growth rate in the northern quahog, Mercenaria mercenaria(Linnaeus, 1758). *Journal of Shellfish Research,* 15**,** 341-344.

CUNNINGHAM, C., HIKIMA, J.-I., JENNY, M., CHAPMAN, R., FANG, G.-C., SASKI, C., LUNDQVIST, M., WING, R., CUPIT, P., GROSS, P., WARR, G. & TOMKINS, J. 2006. New resources for marine genomics: bacterial artificial chromosome libraries for the Eastern and Pacific oysters (Crassostrea virginica and C. gigas). *Marine biotechnology (New York, N.Y.),* 8**,** 521-533.

CUTTER, A. D. 2008. Multilocus Patterns of Polymorphism and Selection Across the X Chromosome of Caenorhabditis remanei. *Genetics,* 178**,** 1661-1672.

CUVIN-ARALAR, M. L. A. & ARALAR, E. V. 1995. Resistance to a heavy metal mixture in Oreochromis Niloticus progenies from parents chronically exposed to the same metals. *Chemosphere,* 30**,** 953-963.

DAKIN, E. & AVISE, J. 2004. Microsatellite null alleles in parentage analysis. *Heredity,* 93**,** 504-509.

DAVEY, G., CAPLICE, N., MARTIN, S. & POWELL, R. 2001. A survey of genes in the Atlantic salmon (*Salmo salar*) as identified by expressed sequence tags. *Gene***,** 121-130.

DAVEY, J., CEZARD, T., FUENTES-UTRILLA, P., ELAND, C., GHARBI, K. & BLAXTER, M. 2013. Special features of RAD Sequencing data: implications for genotyping. *Mol Ecol,* 22**,** 3151-3164.

DAVEY, J., HOHENLOHE, P., ETTER, P., BOONE, J., CATCHEN, J. & BLAXTER, M. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature reviews. Genetics,* 12**,** 499-510.

DAVID, P., PERDIEU, M.-A., PERNOT, A.-F. & JARNE, P. 1997. Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution,* 51**,** 1318-1322.

DAVIDSON, W. S., KOOP, B. F., JONES, S. J., ITURRA, P., VIDAL, R., MAASS, A., JONASSEN, I., LIEN, S. & OMHOLT, S. W. 2010. Sequencing the genome of the Atlantic salmon (Salmo salar). *Genome Biol,* 11**,** 403.

DAVIS, H. C. & CHANLEY, P. E. 1956. Spawning and Egg Production of Oysters and Clams. *Biological Bulletin,* 110**,** 117-128.

DE SANTIS, C., GOMES, G. B. & JERRY, D. R. 2012. Abundance of myostatin gene transcripts and their correlation with muscle hypertrophy during the development of barramundi, Lates calcarifer. *Comp Biochem Physiol B Biochem Mol Biol,* 163**,** 101-107.

DE SANTIS, C., WADE, N. M., JERRY, D. R., PRESTON, N. P., GLENCROSS, B. D. & SELLARS, M. J. 2011. Growing backwards: an inverted role for the shrimp ortholog of vertebrate myostatin and GDF11. *J Exp Biol,* 214**,** 2671-7.

DEL RIO-PORTILLA, M. A. & BEAUMONT 2000. Larval growth, juvenile size and heterozygosity in laboratory reared mussels, Mytilus edulis. *Journal of Experimental Marine Biology and Ecology,* 254**,** 1-17.

DIEHL, W. J. & KOEHN, R. K. 1985. Multiple-locus heterozygosity, mortality, and growth in a cohort of Mytilus edulis. *Marine Biology,* 88**,** 265-271.

DIXON, D. R. 1982. Aneuploidy in mussel embryos (*Mytilus edulis* L.) originating from a polluted dock. *Marine Biology Letters,* 3**,** 155-161.

DIXON, D. R. & FLAVELL, N. 1986. A comparative study of the chromosomes of *Mytilus edulis* and *Mytilus galloprovincialis*. *Journal of the Marine Biological Association of the United Kingdom,* 66**,** 219-228.

DUPERRON, S., BERGIN, C., ZIELINSKI, F., BLAZEJAK, A., PERNTHALER, A., MCKINESS, Z., DECHAINE, E., CAVANAUGH, C. & DUBILIER, N. 2006. A dual symbiosis shared by

two mussel species, Bathymodiolus azoricus and Bathymodiolus puteoserpentis (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environmental microbiology,* 8**,** 1441-1447.

ESTOUP, A., JARNE, P. & CORNUET, J.-M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol,* 11**,** 1591-1604.

EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res,* 8.

FALCONER, D. S. & MACKAY, T. F. C. 1996. *Introduction to Quantitative Genetics*, Longmans Green, Harlow, Essex, UK.

FAO 2010. *The State of World Fisheries and Aquaculture,* Rome.

FAO 2012. *The State of World Fisheries and Aquaculture,* Rome.

FEVOLDEN, S.-E., RØED, K. H. & FJALESTAD, K. T. 2002. Selection response of cortisol and lysozyme in rainbow trout and correlation to growth. *Aquaculture,* 205**,** 61-75.

FJALESTAD, K., MOEN, T. & GOMEZ-RAYA, L. 2003. Prospects for genetic technology in salmon breeding programmes. *Aquaculture Research,* 34**,** 397-406.

FLEURY, E., HUVET, A., LELONG, C., DE LORGERIL, J., BOULO, V., GUEGUEN, Y., BACHÈRE, E., TANGUY, A., MORAGA, D., FABIOUX, C., LINDEQUE, P., SHAW, J., REINHARDT, R., PRUNET, P., DAVEY, G., LAPÈGUE, S., SAUVAGE, C., CORPOREAU, C., MOAL, J., GAVORY, F., WINCKER, P., MOREEWS, F., KLOPP, C., MATHIEU, M., BOUDRY, P. & FAVREL, P. 2009. Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (Crassostrea gigas) assembled into a publicly accessible database: the GigasDatabase. *BMC genomics,* 10**,** 341.

FOLTZ, D. W. 1986. Null Alleles as a Possible Cause of Heterozygote Deficiencies in the Oyster Crassostrea virginica and Other Bivalves. *Evolution,* 40**,** 869-870.

GABILLARD, J.-C., BIGA, P. R., RESCAN, P.-Y. & SEILIEZ, I. 2013. Revisiting the paradigm of myostatin in vertebrates: Insights from fishes. *General and comparative endocrinology,* 194**,** 45-54.

GADGIL, M. & SOLBRIG, O. T. 1972. The Concept of r- and K-Selection: Evidence from Wild Flowers and Some Theoretical Considerations. *The American Naturalist,* 106**,** 14-31.

GAFFNEY, P., PIERCE, J., MACKINLEY, A., TITCHEN, D. & GLENN, W. 2003. Pearl, a novel family of putative transposable elements in bivalve mollusks. *Journal of Molecular Evolution,* 56**,** 308-316.

GAFFNEY, P., SCOTT, T., KOEHN, R. & DIEHL, W. 1990. Interrelationships of heterozygosity, growth rate and heterozygote deficiencies in the coot clam, Mulinia lateralis. *Genetics,* 124**,** 687-699.

GARIKIPATI, D., GAHR, S. & RODGERS, B. 2006. Identification, characterization, and quantitative expression analysis of rainbow trout myostatin-1a and myostatin-1b genes. *J. Endocrinol.,* 190**,** 879-888.

GARRIDO-RAMOS, M. A., STEWART, D. T., SUTHERLAND, B. W. & ZOUROS, E. 1998. The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: Implications for the operation of doubly uniparental inheritance of mitochondrial DNA. *Genome,* 41**,** 818-824.

GARTON, D., KOEHN, R. & SCOTT, T. 1984. Multiple-locus heterozygosity and the physiological energetics of growth in the coot clam, *Mulinia lateralis*, from a natural population. *Genetics,* 108**,** 445-455.

GILBEY, J., VERSPOOR, E., MCLAY, A. & HOULIHAN, D. 2004. A microsatellite linkage map for Atlantic salmon (Salmo salar). *Animal Genetics,* 35**,** 98-105.

GILMOUR, A. R., GOGEL, B. J., CULLIS, B. R., WELHAM, S. J. & THOMPSON, R. 2006. *{ASREML} User Guide Release 2.0*.

GJEDREM, T. & BARANSKI, M. (eds.) 2009. *Selective Breeding in Aquaculture: An Introduction*: Springer.
GJEDREM, T., ROBINSON, N. & RYE, M. 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture,* 350–353**,** 117-129.
GJEDREM, T. & THODESEN, J. 2005. *Selection and Breeding Programs in Aquaculture*, Springer.
GODDARD, M. & HAYES, B. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature reviews. Genetics,* 10**,** 381-391.
GONZÁLEZ-TIZÓN, A., MARTÍNEZ-LAGE, A., REGO, I., AUSIÓ, J. & MÉNDEZ, J. 2000. DNA content, karyotypes, and chromosomal location of 18S-5.8S-28S ribosomal loci in some species of bivalve molluscs from the Pacific Canadian coast. *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada,* 43**,** 1065-1072.
GOSLING, E. 1992. *The mussel Mytilus: ecology, physiology, genetics and culture,* University of California, Elsevier.
GROBET, L., MARTIN, L., PONCELET, D., PIROTTIN, D., BROUWERS, B., RIQUET, J., SCHOEBERLEIN, A., DUNNER, S., MÉNISSIER, F., MASSABANDA, J., FRIES, R., HANSET, R. & GEORGES, M. 1997. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat Genet,* 17**,** 71-74.
GUO, L., LI, L., ZHANG, S., GUO, X. & ZHANG, G. 2011. Novel polymorphisms in the myostatin gene and their association with growth traits in a variety of bay scallop, Argopecten irradians. *Anim Genet,* 42**,** 339-340.
GUTIERREZ, A., LUBIENIECKI, K., DAVIDSON, E., LIEN, S., KENT, M., FUKUI, S., WITHLER, R., SWIFT, B. & DAVIDSON, W. 2012. Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (Salmo salar) using a 6.5K SNP array. *Aquaculture,* 358-359**,** 61-70.
HALL, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* 41**,** 95-98.
HEDGECOCK, D., LI, G., HUBERT, S., BUCKLIN, K. & RIBES, V. 2004. Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster *Crassostrea gigas*. *Journal of Shellfish Research,* 23**,** 379-385.
HILBISH, T. J. & KOEHN, R. 1985. Dominance in physiological phenotypes and fitness at an enzyme locus. *Science,* 229**,** 52-54.
HINEGARDNER, R. 1974. Cellular DNA content of the Mollusca. *Comparative Biochemistry and Physiology Part A: Physiology,* 47**,** 447-460.
HOHENLOHE, P., BASSHAM, S., ETTER, P., STIFFLER, N., JOHNSON, E. & CRESKO, W. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS genetics,* 6.
HOUSTON, R., BISHOP, S., HAMILTON, A., GUY, D., TINCH, A., TAGGART, J., DERAYAT, A., MCANDREW, B. & HALEY, C. 2009. Detection of QTL affecting harvest traits in a commercial Atlantic salmon population. *Anim Genet,* 40**,** 753-755.
HOUSTON, R., DAVEY, J., BISHOP, S., LOWE, N., MOTA-VELASCO, J., HAMILTON, A., GUY, D., TINCH, A., THOMSON, M., BLAXTER, M., GHARBI, K., BRON, J. & TAGGART, J. 2012. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC genomics,* 13**,** 244.
HOUSTON, R., HALEY, C., HAMILTON, A., GUY, D., MOTA-VELASCO, J., GHEYAS, A., TINCH, A., TAGGART, J., BRON, J., STARKEY, W., MCANDREW, B., VERNER-JEFFREYS, D., PALEY, R., RIMMER, G., TEW, I. & BISHOP, S. 2010. The susceptibility of Atlantic

salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity (Edinb),* 105**,** 318-327.

HU, X., GUO, H., HE, Y., WANG, S., ZHANG, L., HUANG, X., ROY, S. W., LU, W., HU, J. & BAO, Z. 2010. Molecular characterization of Myostatin gene from Zhikong scallop Chlamys farreri (Jones et Preston 1904). *Genes Genet Syst,* 85**,** 207-18.

HUBERT, S. & HEDGECOCK, D. 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster Crassostrea gigas. *Genetics,* 168**,** 351-362.

JENNY, M., CHAPMAN, R., MANCIA, A., CHEN, Y., MCKILLEN, D., TRENT, H., LANG, P., ESCOUBAS, J.-M., BACHERE, E., BOULO, V., LIU, Z., GROSS, P., CUNNINGHAM, C., CUPIT, P., TANGUY, A., GUO, X., MORAGA, D., BOUTET, I., HUVET, A., DE GUISE, S., ALMEIDA, J. & WARR, G. 2007. A cDNA microarray for Crassostrea virginica and C. gigas. *Marine biotechnology (New York, N.Y.),* 9**,** 577-591.

JOHNSON, M. S. & BLACK, R. 1984. The wahlund effect and the geographical scale variation in the intertidal limpet *Siphonaria sp. Marine Biology,* 79**,** 295-305.

JONES, D., JERRY, D., FORÊT, S., KONOVALOV, D. & ZENGER, K. 2013. Genome-Wide SNP Validation and Mantle Tissue Transcriptome Analysis in the Silver-Lipped Pearl Oyster, Pinctada maxima. *Marine biotechnology (New York, N.Y.),* 15**,** 647-658.

KAMAL, A. H. M. M. & MAIR, G. C. 2005. Salinity tolerance in superior genotypes of tilapia, Oreochromis niloticus, Oreochromis mossambicus and their hybrids. *Aquaculture,* 247**,** 189-201.

KAMBADUR, R., SHARMA, M., SMITH, T. P. L. & BASS, J. J. 1997. Mutations in myostatin (GDF8) in Double-Muscled Belgian Blue and Piedmontese Cattle. *Genome Research,* 7**,** 910-915.

KENT, M. P., HAYES, B., XIANG, Q., BERG, P. R., GIBBS, R. A. & LIEN, S. 2009. Development of a 6.5K SNP-Chip for Atlantic salmon. *Plant & Animal Genomes XVII Conference.* San Diego, CA, USA.

KERR, T., ROALSON, E. H. & RODGERS, B. D. 2005. Phylogenetic analysis of the myostatin gene sub-family and the differential expression of a novel member in zebrafish. *Evol Dev,* 7**,** 390-400.

KIM, Y.-S., FOX, B., KIM, K. H., LEE, S. B., JIN, H. J. & TAMARU, C. S. 2012. Immersion bath treatment of tilapia fry with myostatin-1 prodomain does not affect tilapia growth at market size. *Aquaculture Research,* 44**,** 1643-1648.

KNIBB, W. R., GORSHKOVA, G. & GORSHKOV, S. 1997. Selection for growth in the gilthead seabream (*Spaurus aurata*). *Israeli Journal of Aquaculture Bamidgeh,* 49**,** 57-66.

KOEHN, R. & GAFFNEY, P. 1984. Genetic heterozygosity and growth rate in Mytilus edulis. *Marine Biology,* 82**,** 1-7.

KOEHN, R. K., MILKMAN, R. & MITTON, J. B. 1976. Population Genetics of Marine Pelecypods. IV. Selection, Migration and Genetic Differentiation in the Blue Mussel Mytilus edulis. *Evolution,* 30**,** 2-32.

KONG, N., LI, Q., YU, H. & KONG, L.-F. 2013. Heritability estimates for growth-related traits in the Pacific oyster (Crassostrea gigas) using a molecular pedigree. *Aquaculture Research***,** 1-10.

KOOP, B., VON SCHALBURG, K., LEONG, J., WALKER, N., LIEPH, R., COOPER, G., ROBB, A., BEETZ-SARGENT, M., HOLT, R., MOORE, R., BRAHMBHATT, S., ROSNER, J., REXROAD, C., MCGOWAN, C. & DAVIDSON, W. 2008. A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics,* 9**,** 545.

KOURTIDIS, A., DROSOPOULOU, E., N. PANTZARTZI, C., C. CHINTIROGLOU, C. & G. SCOURAS, Z. 2006. Three new satellite sequences and a mobile element found inside HSP70 introns of the Mediterranean mussel (Mytilus galloprovincialis). *Genome,* 49**,** 1451-1458.

KRUGLYAK, L. & NICKERSON, D. 2001. Variation is the spice of life. *Nature Genetics,* 27**,** 234-236.

KÜLHEIM, C., YEOH, S., MAINTZ, J., FOLEY, W. & MORAN, G. 2009. Comparative SNP diversity among four Eucalyptus species for genes from secondary metabolite biosynthetic pathways. *BMC genomics,* 10**,** 452.

KUMAR, S., BANKS, T. & CLOUTIER, S. 2012. SNP Discovery through Next-Generation Sequencing and Its Applications. *International Journal of Plant Genomics,* 2012.

KWOK, S., KELLOGG, D. E., MCKINNEY, N., SPASIC, D., GODA, L., LEVENSON, C. & SNINSKY, J. J. 1990. Effects of primer-template mismtach on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Research,* 18**,** 999-1005.

LALLIAS, D., LAPÈGUE, S., HECQUET, C., BOUDRY, P. & BEAUMONT, A. R. 2007. AFLP-based genetic linkage maps of the blue mussel (Mytilus edulis). *Animal Genetics,* 38**,** 340–349.

LAMB, B. C. 2000. *The applied genetics of plants, animals, humans and fungi*, Imperial College Press.

LANDEGREN, U., NILSSON, M. & KWOK, P. 1998. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res,* 8**,** 769-776.

LANGDON, C., EVANS, F., JACOBSON, D. & BLOUIN, M. 2003. Yields of cultured Pacific oysters Crassostrea gigas Thunberg improved after one generation of selection. *Aquaculture,* 220**,** 227-244.

LAUNEY, S. & HEDGECOCK, D. 2001. High genetic load in the Pacific oyster Crassostrea gigas. *Genetics,* 159**,** 255-265.

LAWSON, D., HELLENTHAL, G., MYERS, S. & FALUSH, D. 2012. Inference of population structure using dense haplotype data. *PLoS genetics,* 8.

LEE, C. Y., HU, S. Y., GONG, H. Y., CHEN, M. H., LU, J. K. & WU, J. L. 2009. Suppression of myostatin with vector-based RNA interference causes a double-muscle effect in transgenic zebrafish. *Biochem Biophys Res Commun,* 387**,** 766-71.

LEE, S. B., KIM, Y. S., OH, M.-Y., JEONG, I.-H., SEONG, K.-B. & JIN, H.-J. 2010. Improving rainbow trout (Oncorhynchus mykiss) growth by treatment with a fish (Paralichthys olivaceus) myostatin prodomain expressed in soluble forms in E. coli. *Aquaculture,* 302**,** 270-278.

LEE, S. J. & MCPHERRON, A. C. 1999. Myostatin and the control of skeletal muscle mass. *Curr Opin Genet Dev,* 9**,** 604-7.

LI, G., HUBERT, S., BUCKLIN, K., RIBES, V. & HEDGECOCK, D. 2003. Characterization of 79 microsatellite DNA markers in the Pacific oyster Crassostrea gigas. *Molecular Ecology Notes,* 3**,** 228-232.

LI, H., FAN, J., LIU, S., YANG, Q., MU, G. & HE, C. 2012. Characterization of a myostatin gene (MSTN1) from spotted halibut (Verasper variegatus) and association between its promoter polymorphism and individual growth performance. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology,* 161**,** 315-322.

LI, Q., WANG, Q., LIU, S. & KONG, L. 2011. Selection response and realized heritability for growth in three stocks of the Pacific oyster Crassostrea gigas. *Fisheries Science,* 77**,** 643-648.

LIEN, S., GIDSKEHAUG, L., MOEN, T., HAYES, B., BERG, P., DAVIDSON, W., OMHOLT, S. & KENT, M. 2011. A dense SNP-based linkage map for Atlantic salmon (Salmo salar) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC genomics,* 12**,** 615.

LIN, L., YINHU, L., SILIANG, L., NI, H., YIMIN, H., RAY, P., DANNI, L., LIHUA, L. & MAGGIE, L. 2012. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology,* 2012.

LIU, L., YU, X. & TONG, J. 2012. Molecular characterization of myostatin (MSTN) gene and association analysis with growth traits in the bighead carp (Aristichthys nobilis). *Mol Biol Rep,* 39**,** 9211-21.

MACAVOY, E. S., WOOD, A. R. & GARDNER, J. P. A. 2008. Development and evaluation of microsatellite markers for identification of individual Greenshell™ mussels (Perna canaliculus) in a selective breeding programme. *Aquaculture,* 274**,** 41-48.

MACCATROZZO, L., BARGELLONI, L., CARDAZZO, B., RIZZO, G. & PATARNELLO, T. 2001. A novel second myostatin gene is present in teleost fish. *FEBS Lett,* 509**,** 36-40.

MALLET, A. L., ZOUROS, E., GARTNER-KEPKAY, K. E., FREEMAN, K. R. & DICKIE, L. M. 1985. Larval viability and heterozygote deficiency in populations of marine bivalves: evidence from pair matings of mussels. *Marine Biology,* 87**,** 165-172.

MARTÍNEZ-LAGE, A., GONZÁLEZ-TIZÓN, A., AUSIÓ, J. & MÉNDEZ, J. 1997. Karyotypes and Ag-NORs of the mussels mytilus californianus and M. trossulus from the Pacific Canadian coast. *Aquaculture,* 153**,** 239-249.

MCPHERRON, A. C., LAWLER, A. M. & LEE, S. J. 1997. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature,* 387**,** 83-90.

MCPHERRON, A. C. & LEE, S. J. 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proc Natl Acad Sci U S A,* 94**,** 12457-61.

MENGNING MAUREEN, L., JOHN, W. D., RUBY, B., JIE, H., FENGTANG, Y., AZIZ, A., MARK, L. B., ANGUS, D. & ZHANJIANG, L. 2013. Fine Mapping of the Pond Snail Left-Right Asymmetry (Chirality) Locus Using RAD-Seq and Fibre-FISH. *PLoS One,* 8.

MILLER, C., JOYCE, P. & WAITS, L. 2002. Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics,* 160**,** 357-366.

MOEN, T., BARANSKI, M., SONESSON, A. & KJØGLUM, S. 2009. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. *BMC genomics,* 10**,** 368.

MOEN, T., HAYES, B., BARANSKI, M., BERG, P., KJØGLUM, S., KOOP, B., DAVIDSON, W., OMHOLT, S. & LIEN, S. 2008. A linkage map of the Atlantic salmon (Salmo salar) based on EST-derived SNP markers. *BMC genomics,* 9**,** 223.

MORIN, P. A., CHAMBERS, K. E., BOESCH, C. & VIGILANT, L. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (Pan troglodytes verus). *Molecular Ecology,* 10**,** 1835-1844.

MUIR, J. 2005. Managing to harvest? Perspectives on the potential of aquaculture. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 360**,** 191-218.

NEELY, K. G., MYERS, J. M., HARD, J. J. & SHEARER, K. D. 2008. Comparison of growth, feed intake, and nutrient efficiency in a selected strain of coho salmon (Oncorhynchus kisutch) and its source stock. *Aquaculture,* 283**,** 134-140.

NELL, J., SMITH, I. & SHERIDAN, A. 1999. Third generation evaluation of Sydney rock oyster *Saccostrea commercialis* (Iredale and Roughley) breeding lines. *Aquaculture,* 37**,** 693-700.

NEWELL, R. 2004. Ecosystem influences of natural and cultivated populations of suspension-feeding bivalve molluscs: a review. *Journal of Shellfish Research,* 23**,** 51-61.

NEWKIRK, G. F. & HALEY, L. E. 1983. Selection for growth rate in the European oyster, Ostrea edulis: Response of second generation groups. *Aquaculture,* 33**,** 149-155.

NGUYEN, T. T. T., HAYES, B. J., GUTHRIDGE, K., AB RAHIM, E. S. & INGRAM, B. A. 2011. Use of a microsatellite-based pedigree in estimation of heritabilities for economic traits in Australian blue mussel, Mytilus galloprovincialis. *Journal of Animal Breeding and Genetics,* 128**,** 482-490.

ØDEGÅRD, J., OLESEN, I., GJERDE, B. & KLEMETSDAL, G. 2007. Positive genetic correlation between resistance to bacterial (furunculosis) and viral (infectious salmon anaemia) diseases in farmed Atlantic salmon (Salmo salar). *Aquaculture,* 271**,** 173-177.

OSTBYE, T.-K., WETTEN, O., TOOMING-KLUNDERUD, A., JAKOBSEN, K., YAFE, A., ETZIONI, S., MOEN, T. & ANDERSEN, O. 2007. Myostatin (MSTN) gene duplications in Atlantic salmon (Salmo salar): evidence for different selective pressure on teleost MSTN-1 and -2. *Gene,* 403**,** 159-169.

PANOVA, M., MAKINEN, T., FOKIN, M., ANDRE, C. & JOHANNESSON, K. 2008. Microsatellite cross-species amplification in the genus Littorina and detection of null alleles in Littorina saxatilis. *Journal of Molluscan Studies,* 74**,** 111-117.

PANTE, E., ROHFRITSCH, A., BECQUET, V., BELKHIR, K., BIERNE, N. & GARCIA, P. 2012. SNP detection from de novo transcriptome sequencing in the bivalve Macoma balthica: marker development for evolutionary studies. *PLoS One,* 7.

PECHENIK, J. A. 1999. On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology Progress Series,* 177**,** 269-297.

POGSON, G. 1991. Expression of overdominance for specific activity at the phosphoglucomutase-2 locus in the Pacific oyster, Crassostrea gigas. *Genetics,* 128**,** 133-141.

POGSON, G. & ZOUROS, E. 1994. Allozyme and RFLP heterozygosities as correlates of growth rate in the scallop Placopecten magellanicus: a test of the associative overdominance hypothesis. *Genetics,* 137**,** 221-231.

POLATO, N., VERA, J. & BAUMS, I. 2011. Gene discovery in the threatened elkhorn coral: 454 sequencing of the Acropora palmata transcriptome. *PLoS One,* 6.

PONZONI, R., ACOSTA, B. & PONNIAH, A. (eds.) 2006. *Development of Aquatic Animal Genetic Improvement and Dissemination Programs: Current Status and Action Plans*: WorldFish Center Conference Proceedings

POSTLETHWAIT, J., YAN, Y., GATES, M., HORNE, S., AMORES, A., BROWNLIE, A., DONOVAN, A., EGAN, E., FORCE, A., GONG, Z., GOUTEL, C., FRITZ, A., KELSH, R., KNAPIK, E., LIAO, E., PAW, B., RANSOM, D., SINGER, A., THOMSON, M., ABDULJABBAR, T., YELICK, P., BEIER, D., JOLY, J., LARHAMMAR, D., ROSA, F., WESTERFIELD, M., ZON, L., JOHNSON, S. & TALBOT, W. 1998. Vertebrate genome evolution and the zebrafish gene map. *Nat Genet,* 18**,** 345-349.

POWELL, J., WHITE, I., GUY, D. & BROTHERSTONE, S. 2008. Genetic parameters of production traits in Atlantic salmon (Salmo salar). *Aquaculture,* 274**,** 225-231.

QUAIL, M., SMITH, M., COUPLAND, P., OTTO, T., HARRIS, S., CONNOR, T., BERTONI, A., SWERDLOW, H. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics,* 13**,** 341.

RAWSON, P. D. & HILBISH, T. J. 1990. Heritability of juvenile growth for the hard clam Mercenaria mercenaria. *Marine Biology,* 105**,** 429-436.

REECE, K., RIBEIRO, W., GAFFNEY, P., CARNEGIE, R. & ALLEN, S. 2004. Microsatellite marker development and analysis in the eastern oyster (Crassostrea virginica): confirmation of null alleles and non-Mendelian segregation ratios. *J Hered,* 95**,** 346-352.

REES, G., POND, K., KAY, D., BARTRAM, J. & SANTO DOMINGO, J. (eds.) 2010. *Safe Management of Shellfish and Harvest Waters,* London, UK: IWA.

REID, D., SZANTO, A., GLEBE, B., DANZMANN, R. & FERGUSON, M. 2005. QTL for body weight and condition factor in Atlantic salmon (Salmo salar): comparative analysis with rainbow trout (Oncorhynchus mykiss) and Arctic charr (Salvelinus alpinus). *Heredity (Edinb),* 94**,** 166-172.

RICE, P., LONGDEN, I. & BLEASBY, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG,* 16**,** 276-277.

RICKE, R. M. & DEURSEN, J. M. V. 2013. Aneuploidy in health, disease, and aging. *J Cell Biol,* 201**,** 11-21.

RISE, M. L. 2004. Development and Application of a Salmonid EST Database and cDNA Microarray: Data Mining and Interspecific Hybridization Characteristics. *Genome Research,* 14**,** 478-490.

RODRIGUEZ-JUIZ, A. M., TORRADO, M. & MENDEZ, J. 1996. Genome-size variation in bivalve molluscs determined by flow cytometry. *Marine Biology,* 126**,** 489-497.

ROWE, H., RENAUT, S. & GUGGISBERG, A. 2011. RAD in the realm of next-generation sequencing technologies. *Mol Ecol,* 20**,** 3499-3502.

RUBIN, B., REE, R. & MOREAU, C. 2012. Inferring phylogenies from RAD sequence data. *PLoS One,* 7.

SAAVEDRA, C. & BACHÈRE, E. 2006. Bivalve genomics. *Aquaculture,* 256**,** 1-14.

SAAVEDRA, C., ZAPATA, C., GUERRA, A. & ALVAREZ, G. 1993. Allozyme variation in European populations of the oyster Ostrea edulis. *Marine Biology,* 115**,** 85-95.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S., KAKOL, J., STEIN, L., MARTH, G., SHERRY, S., MULLIKIN, J., MORTIMORE, B., WILLEY, D., HUNT, S., COLE, C., COGGILL, P., RICE, C., NING, Z., ROGERS, J., BENTLEY, D., KWOK, P., MARDIS, E., YEH, R., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L., HILLIER, L., WATERSTON, R., MCPHERSON, J., GILMAN, B., SCHAFFNER, S., VAN ETTEN, W., REICH, D., HIGGINS, J., DALY, M., BLUMENSTIEL, B., BALDWIN, J., STANGE-THOMANN, N., ZODY, M., LINTON, L., LANDER, E., ALTSHULER, D. & INTERNATIONAL, S. N. P. M. W. G. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-933.

SANCHEZ-RAMOS, I., CROSS, I., MACHA, J., MARTINEZ-RODRIGUEZ, G., KRYLOV, V. & REBORDINOS, L. 2012. Assessment of tools for marker-assisted selection in a marine commercial species: significant association between MSTN-1 gene polymorphism and growth traits. *ScientificWorldJournal,* 2012**,** 369802.

SANGER, F., NICKLEN, S. & COULSON, A. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-5467.

SARVER, S. K., KATOH, M. & FOLTZ, D. W. 1992. Apparent overdominance of enzyme specific activity in two marine bivalves. *Genetica,* 85**,** 231-239.

SAUVAGE, C., BIERNE, N., LAPÈGUE, S. & BOUDRY, P. 2007. Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster Crassostrea gigas. *Gene,* 406**,** 13-22.

SAWATARI, E., SEKI, R., ADACHI, T., HASHIMOTO, H., UJI, S., WAKAMATSU, Y., NAKATA, T. & KINOSHITA, M. 2010. Overexpression of the dominant-negative form of myostatin results in doubling of muscle-fiber number in transgenic medaka (Oryzias latipes). *Comparative Biochemistry and Physiology Part A Molecular & Integrative Physiology,* 155**,** 183-9.

SERNAPESCA 2011. Anuario Estadistico de Pesca. Chile.

SIITONEN, L. & GALL, G. A. E. 1989. Response to selection for early spawn date in rainbow trout, Salmo gairdneri. *Aquaculture,* 78**,** 153-161.

SINGER, V., JONES, L., YUE, S. & HAUGLAND, R. 1997. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. *Anal Biochem,* 249**,** 228-238.

SINGH, S. M. 1982. Enzyme heterozygosity associated with grwoth at different developmental stages in oysters. *Canadian Journal of Genetics and Cytology,* 24**,** 451-458.

SINGH, S. M. & ZOUROS, E. 1978. Genetic Variation Associated with Growth Rate in the American Oyster (Crassostrea virginica). *Evolution,* 32**,** 342-353.

SMALL, K., BRUDNO, M., HILL, M. & SIDOW, A. 2007. Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 5698-5703.

SPRUNG, M. 1984. Physiological energetics of mussel larvae (*Mytilus edulis*). II. Food uptake. *Marine Ecology Progress Series,* 17**,** 295-305.

STEPHENS, M. & DONNELLY, P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American journal of human genetics,* 73**,** 1162-1169.

STEPHENS, M., SMITH, N. & DONNELLY, P. 2001. A new statistical method for haplotype reconstruction from population data. *American journal of human genetics,* 68**,** 978-989.

STINCKENS, A., GEORGES, M. & BUYS, N. 2011. Mutations in the myostatin gene leading to hypermuscularity in mammals: indications for a similar mechanism in fish? *Anim Genet,* 42**,** 229-34.

STRACHAN, T. & READ, A. P. 1999. Instability of the human genome: mutation and DNA repair. *Human Molecular Genetics.* New York: Willey-Liss.

STRÖMGREN, T. & NIELSEN, M. V. 1989. Heritability of growth in larvae and juveniles of Mytilus edulis. *Aquaculture,* 80**,** 1-6.

SUN, Y., YU, X. & TONG, J. 2012. Polymorphisms in Myostatin Gene and Associations with Growth Traits in the Common Carp (Cyprinus carpio L.). *Int J Mol Sci,* 13**,** 14956-61.

TANGUY, A., BIERNE, N., SAAVEDRA, C., PINA, B., BACHÈRE, E., KUBE, M., BAZIN, E., BONHOMME, F., BOUDRY, P., BOULO, V., BOUTET, I., CANCELA, L., DOSSAT, C., FAVREL, P., HUVET, A., JARQUE, S., JOLLIVET, D., KLAGES, S., LAPÈGUE, S., LEITE, R., MOAL, J., MORAGA, D., REINHARDT, R., SAMAIN, J.-F., ZOUROS, E. & CANARIO, A. 2008. Increasing genomic information in bivalves through new EST collections in four species: development of new genetic markers for environmental studies and genome evolution. *Gene,* 408**,** 27-36.

TAYLOR, J. S., VAN DE PEER, Y. & MEYER, A. 2001. Genome duplication, divergent resolution and speciation. *Trends in Genetics,* 17**,** 299-301.

TEAM, R. D. C. 2011. R: A language and environment for statitical computing. Vienna, Austria: R Foundation for Statitical Computing.

THEOLOGIDIS, I., FODELIANAKIS, S., GASPAR, M. B. & ZOUROS, E. 2008. Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA in Donax Trunculus (Bivalvia: Donacidae) and the Problem of Its Sporadic Detection in Bivalvia. *Evolution,* 62**,** 959-970.

THIRIOT-QUIEVREUX, C. 2002. Review of the literature on bivalve cytogenetics in the last ten years. *Cahiers de Biologie Marine*.

THIRIOT-QUIÉVREUX, C. & INSUA, A. 1992. Nucleolar organiser region variation in the chromosomes of three oyster species. *Journal of Experimental Marine Biology and Ecology***,** 33-40.

THOMPSON, R. J. 1979. Fecundity and Reproductive Effort in the Blue Mussel (Mytilus edulis), the Sea Urchin (Strongylocentrotus droebachiensis), and the Snow Crab

(Chionoecetes opilio) from Populations in Nova Scotia and Newfoundland. *Journal of the Fisheries Research Board of Canada,* 36**,** 955-964.

THORGAARD, G., BAILEY, G., WILLIAMS, D., BUHLER, D., KAATTARI, S., RISTOW, S., HANSEN, J., WINTON, J., BARTHOLOMEW, J., NAGLER, J., WALSH, P., VIJAYAN, M., DEVLIN, R., HARDY, R., OVERTURF, K., YOUNG, W., ROBISON, B., REXROAD, C. & PALTI, Y. 2002. Status and opportunities for genomics research with rainbow trout. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology,* 133**,** 609-646.

THORSEN, J., ZHU, B., FRENGEN, E., OSOEGAWA, K., DE JONG, P., KOOP, B., DAVIDSON, W. & HØYHEIM, B. 2005. A highly redundant BAC library of Atlantic salmon (Salmo salar): an important tool for salmon projects. *BMC Genomics,* 6**,** 50.

THORSON, G. 1950. REPRODUCTIVE and LARVAL ECOLOGY OF MARINE BOTTOM INVERTEBRATES. *Biological Reviews,* 25**,** 1-45.

THORVALDSDÓTTIR, H., ROBINSON, J. & MESIROV, J. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics,* 14**,** 178-192.

TORO, J., ALCAPAN, A., OJEDA, J. & VERGARA, A. M. 2004. Selection response for growth rate (shell height and live weight) in the Chilean blue mussel (Mytilus chilensis Hupe 1854). *Journal of Shellfish Research,* 23**,** 753-757.

TORO, J. E. & PAREDES, L. I. 1996. Heritability estimates of larval shell length in the Chilean blue mussel Mytilus chilensis, under different food densities. *Aquatic Living Resources,* 9**,** 347-350.

TRACEY, M. L., BELLET, N. F. & GRAVEM, C. D. 1975. Excess allozyme homozygosity and breeding population structure in the Mussel Mytilus californianus. *Marine Biology,* 32**,** 303-311.

URIARTE, I. 2008. Estado actual del cultivo de moluscos bivalvos en Chile *In:* LOVATELLI, A., FARIAS, A. & URIARTE, I. (eds.) *Estado actual del cultivo y manejo de moluscos bivalvos y su proyeccion futura: factores que afectan su sustentabilidad en America Latina.* Roma: FAO.

VALENZUELA-MUÑOZ, V., ARAYA-GARAY, J. M. & GALLARDO-ESCÁRATE, C. 2013. SNP discovery and High Resolution Melting Analysis from massive transcriptome sequencing in the California red abalone Haliotis rufescens. *Marine Genomics,* 10**,** 11-16.

VENIER, P., DE PITTÀ, C., BERNANTE, F., VAROTTO, L., DE NARDI, B., BOVO, G., ROCH, P., NOVOA, B., FIGUERAS, A., PALLAVICINI, A. & LANFRANCHI, G. 2009. MytiBase: a knowledgebase of mussel (M. galloprovincialis) transcribed sequences. *BMC genomics,* 10**,** 72.

WADA, K. T. 1986. Genetic selection for shell traits in the Japanese pearl oyster, Pinctada fucata martensii. *Aquaculture,* 57**,** 171-176.

WAHLUND, S. 1928. ZUSAMMENSETZUNG VON POPULATIONEN UND KORRELATIONSERSCHEINUNGEN VOM STANDPUNKT DER VERERBUNGSLEHRE AUS BETRACHTET. *Hereditas,* 11**,** 65-106.

WALDMAN, B. & MCKINNON, J. S. 1993. Inbreeding and outbreeding in fishes, amphibians and reptiles. *The natural history of inbreeding and outbreeding: theoretical and empirical perspectives. University of Chicago Press, Chicago***,** 250-282.

WANG, D., FAN, J., SIAO, C., BERNO, A., YOUNG, P., SAPOLSKY, R., GHANDOUR, G., PERKINS, N., WINCHESTER, E., SPENCER, J., KRUGLYAK, L., STEIN, L., HSIE, L., TOPALOGLOU, T., HUBBELL, E., ROBINSON, E., MITTMANN, M., MORRIS, M., SHEN, N., KILBURN, D., RIOUX, J., NUSBAUM, C., ROZEN, S., HUDSON, T., LIPSHUTZ, R., CHEE, M. & LANDER, E. 1998. Large-scale identification, mapping, and genotyping of single-

nucleotide polymorphisms in the human genome. *Science (New York, N.Y.),* 280**,** 1077-1082.

WANG, Q., LI, Q., KONG, L. & YU, R. 2012. Response to selection for fast growth in the second generation of Pacific oyster (Crassostrea gigas). *Journal of Ocean University of China,* 11**,** 413-418.

WASSER, S. K., MAILAND, C., BOOTH, R., MUTAYOBA, B., KISAMO, E., CLARK, B. & STEPHENS, M. 2007. Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proceedings of the National Academy of Sciences,* 104**,** 4228-4233.

WATSON, J. D. & CRICK, F. H. 1953. Molecular structure of nucleic acids. *Nature,* 171**,** 737-738.

WATTIER, R., ENGEL, C. R., SAUMITOU-LAPRADE, P. & VALERO, M. 1998. Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in Gracilaria gracilis (Rhodophyta). *Molecular Ecology,* 7**,** 1569-1573.

WILLING, E. M., HOFFMANN, M., KLEIN, J. D., WEIGEL, D. & DREYER, C. 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics,* 27**,** 2187-2193.

WOODS, I., KELLY, P., CHU, F., NGO-HAZELETT, P., YAN, Y. L., HUANG, H., POSTLETHWAIT, J. & TALBOT, W. 2000. A comparative map of the zebrafish genome. *Genome Research,* 10**,** 1903-1914.

XING, F., TAN, X., ZHANG, P. J., MA, J., ZHANG, Y., XU, P. & XU, Y. 2007. Characterization of amphioxus GDF8/11 gene, an archetype of vertebrate MSTN and GDF11. *Dev Genes Evol,* 217**,** 549-54.

YAOHUA, S., HONG, K., XIMING, G., ZHIFENG, G., YAN, W. & AIMIN, W. 2009. Genetic linkage map of the pearl oyster, Pinctada martensii (Dunker). *Aquaculture Research,* 41**,** 35-44.

YU, Z. & GUO, X. 2003. Genetic linkage map of the eastern oyster Crassostrea virginica Gmelin. *The Biological bulletin,* 204**,** 327-338.

YUND, P. O. & NEIL, P. G. O. 2000. Microgeographic genetic differentiation in a colonial ascidian (Botryllus schlosseri) population. *Marine Biology,* 137**,** 583-588.

ZHANG, G., FANG, X., GUO, X., LI, L., LUO, R., XU, F., YANG, P., ZHANG, L., WANG, X., QI, H., XIONG, Z., QUE, H., XIE, Y., HOLLAND, P., PAPS, J., ZHU, Y., WU, F., CHEN, Y., WANG, J., PENG, C., MENG, J., YANG, L., LIU, J., WEN, B., ZHANG, N., HUANG, Z., ZHU, Q., FENG, Y., MOUNT, A., HEDGECOCK, D., XU, Z., LIU, Y., DOMAZET-LOŠO, T., DU, Y., SUN, X., ZHANG, S., LIU, B., CHENG, P., JIANG, X., LI, J., FAN, D., WANG, W., FU, W., WANG, T., WANG, B., ZHANG, J., PENG, Z., LI, Y., LI, N., WANG, J., CHEN, M., HE, Y., TAN, F., SONG, X., ZHENG, Q., HUANG, R., YANG, H., DU, X., CHEN, L., YANG, M., GAFFNEY, P., WANG, S., LUO, L., SHE, Z., MING, Y., HUANG, W., ZHANG, S., HUANG, B., ZHANG, Y., QU, T., NI, P., MIAO, G., WANG, J., WANG, Q., STEINBERG, C., WANG, H., LI, N., QIAN, L., ZHANG, G., LI, Y., YANG, H., LIU, X., WANG, J., YIN, Y. & WANG, J. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature,* 490**,** 49-54.

ZHANG, L. & GUO, X. 2010. Development and validation of single nucleotide polymorphism markers in the eastern oyster Crassostrea virginica Gmelin by mining ESTs and resequencing. *Aquaculture,* 302**,** 124-129.

ZHU, Y. Y., LIANG, H. W., LI, Z., LUO, X. Z., LI, L., ZHANG, Z. W. & ZOU, G. W. 2012. [Polymorphism of MSTN gene and its association with growth traits in yellow catfish (Pelteobagruse fulvidraco)]. *Yi Chuan,* 34**,** 72-8.

ZOUROS, E. & FOLTZ, D. W. 1983. Minimal selection requirements for the correlation between heterozygosity and growth, and for the deficiency of heterozygotes, in oyster population. *Developmental Genetics,* 4**,** 393-405.

ZOUROS, E. & FOLTZ, D. W. 1984. Possible explanations of heterozygote deficiency in bivalve molluscs. *Malacologia,* 25**,** 583-591.

ZOUROS, E., OBERHAUSER, A., SAAVEDRA, C. & FREEMAN, K. R. 1994. An unusual type of mitochondiral DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 7463-7467.

ZOUROS, E., ROMERO-DOREY, M. & MALLET, A. L. 1988. Heterozygosity and Growth in Marine Bivalves: Further Data and Possible Explanations. *Evolution,* 42**,** 1332-1341.

ZOUROS, E., SINGH, S. M., FOLTZ, D. W. & MALLET, A. L. 1983. Post-settlement viability in the American oyster (Crassostrea virginica): an overdominant phenotype. *Genetics Research,* 41**,** 259-270.

ZOUROS, E., SINGH, S. M. & MILES, H. E. 1980. Growth Rate in Oysters: An Overdominant Phenotype and Its Possible Explanations. *Evolution,* 34**,** 856-867.